The genomics, evolution, and spread of *Shigella* species in Asia



Chung The Hao

Kellogg College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2018

The genomics, evolution, and spread of *Shigella* spp. in Asia

Chung The Hao, Kellogg College, DPhil. in Clinical Medicine, Trinity Term 2018

Abstract

The bacterial genus *Shigella* remains a significant cause of diarrhoea in young children, particularly in tropical and developing countries. Effective clinical management and public health policy implementation require an in-depth understanding of the pathogen's evolution and epidemiology. This thesis set out to utilize genomic and phylogenetic analyses to investigate *Shigella*'s life history in several settings. Firstly, I focused on *S. sonnei* exhibiting escalating resistance to fluoroquinolones, which are the recommended treatment for shigellosis. I showed that extant geographically dispersed fluoroquinolone resistant *S. sonnei* emerged as a single clone in South Asia around 2007. This clone then spread internationally, creating sustained transmissions in Southeast Asia and Europe. Secondly, examination of a vast historical *Shigella* sequencing dataset from Southeast Asia showed that contemporary regional *Shigella* populations are progenies of relatively recent introductions, mostly during the 1980s-1990s. Intraspecific clonal replacement, occurring in both *S. sonnei* and major *S. flexneri* serotypes, was frequent and linked with resistance to contemporaneously used antimicrobials. The secondary aim of this thesis was to explore the alternations of the gut microbiota in response to infectious diarrhoea in children. I found that the diarrhoeal faecal microbiome could either generally resemble or greatly diverge from that of a healthy state. *Escherichia* and oral bacterial commensals were found to be consistently enriched in the diarrhoeal faecal microbiome. These studies demonstrate the versatility of applying sequencing to studying diarrhoea in endemic regions, enhancing our current understanding of this public health problem.

Acknowledgements

My utmost gratitude goes to my supervisor, Prof. Stephen Baker, who foremost has given me a fantastic opportunity to conduct research in Vietnam. I particularly thank him for his unwavering guidance, support, and trust. I owe many thanks to Prof. Nicholas Thomson, who has offered me access to excellent genomic sequencing and analysis capacity, as well as many thoughtful advices. I am also grateful to Dr. Maia Rabaa for countless hours of dedicated help and insightful conversations on my thesis and career. I must extend my heartfelt thanks to all members of the Enteric Infections group, particularly Ms. Chau, Ms. Nhi, Ms. Nhu, Ms. Trang, Ms. Duong, Ms. Phuc, Mr. Trung, Mr. Phat, Mr. Duy, Ms. Tam, Ms. To, and Ms. Anh. Without them, my study would have been much more challenging and much less enjoyable. Much appreciation also goes to my dearest friends, especially Mr. Chanh Truong and Ms. Jacqueline Leung, who have helped me cope with the difficult times throughout these years.

I am indebted to the diligence and generosity of multiple collaborators, whose contributions have made this thesis possible. These include but are not limited to the Wellcome Trust Sanger Institute – team 216, AFRIMS Thailand, Public Health England, Pasteur Institute in Paris, University of Melbourne, and Genome Institute of Singapore. I am also thankful to the Centre of Tropical Diseases in Oxford for funding my study.

Finally, I thank Mom, Dad, and my little brother Luan for their immense trust, patience and support.

Declaration

I, Chung The Hao, confirm that the work presented in this thesis is my own and was conducted at the Oxford University Clinical Research Unit in Ho Chi Minh city, Vietnam. Where information or work has been derived from other sources, I confirm that this has been indicated in the thesis. This thesis has not been submitted for a degree or other qualification to this or any other university.

"*Nothing in Biology Makes Sense Except in the Light of Evolution.*"

Theodosius Dobzhansky (1973)

Table of Contents

Abstract ..... **i**

Acknowledgements ..... **ii**

Declaration ..... **iii**

Table of contents ..... **v**

List of tables ..... **xi**

List of figures ..... **xiii**

Abbreviations ..... **xvii**

Publications during DPhil candidature ..... **xxii**

1. Introduction **1**

1.1 Global burden of diarrhoeal diseases 1

1.2 The diarrhoeal gut microbiota 2

1.2.1 The human gut microbiota and its development during infancy 2

1.2.2 Impact of diarrhoea on gut microbiota 5

1.3 The bacterial genus *Shigella* 9

1.3.1 Nomenclature and classification 9

1.3.2 Laboratory identification of *Shigella* 10

1.3.3 Molecular mechanism of *Shigella* pathogenesis 11

1.3.4 Epidemiology of *Shigella* 15

1.4 Shigellosis treatment and management 19

1.4.1 Treatment of shigellosis and antimicrobial resistance 19

1.4.2 Progress in *Shigella* vaccines 21

1.5 Evolutionary history of the genus *Shigella* 22

1.5.1 The phylogenetic relationships amongst *Shigella* spp. 22
1.5.2 Evolutionary history of the *Shigella* virulence plasmid 28
1.5.3 Acquisition of genetic elements 31
1.5.4 Reductive evolution 32
1.6 Genomic evolution of *Shigella sonnei* 34
1.6.1 Global population structure of *Shigella sonnei* 34
1.6.2 Evolution of *Shigella sonnei* in Vietnam 37
1.7 Genomic evolution of *Shigella flexneri* 39
1.7.1 Genomic determinants of *Shigella flexneri* serotypes 39
1.7.2 Global population structure of *Shigella flexneri* 42
1.8 Genomic evolution of *Shigella dysenteriae* and *Shigella boydii* 46
1.9 Aims and objectives of the thesis 48
2. Materials and Methods 49
2.1 Methods applicable throughout chapters 3 to 5 49
2.1.1 Whole genome sequencing (WGS) of *Shigella* isolates 49
2.1.2 Mapping sequencing reads to references and detection of single nucleotide polymorphisms (SNPs) 49
2.1.3 Phylogenetic reconstructions 50
2.1.4 Temporal analysis and Bayesian phylogenetic inference 51
2.1.5 Construction of de novo assemblies 53
2.1.6 Determination of the accessory genome 54
2.1.7 Stochastic mapping of geographical traits 54

2.2 Methodology specific to Chapter 3: The introduction and establishment of fluoroquinolone resistant *Shigella sonnei* into Bhutan 55
2.2.1 Sample collection 55
2.2.2 Bacteriology and antimicrobial susceptibility testing 56
2.2.3 Whole genome sequencing 56
2.2.4 Phylogenetic reconstruction of Bhutanese *Shigella sonnei* 57
2.2.5 Gene content analysis 58
2.3 Methodology specific to Chapter 4: The emergence and global spread of fluoroquinolone resistant *Shigella sonnei* 59
2.3.1 Organism collection and whole genome sequencing 59
2.3.2 Phylogenetic analysis 59
2.3.3 Hierarchical Bayesian clustering 60
2.3.4 Bayesian phylogenetic inference 60
2.3.5 The geographical spread of Central Asia III *Shigella sonnei* 61
2.3.6 Analysis on pseudogenes and gene loss 62
2.3.7 Substitution mutation analysis 63
2.3.8 Cytokine profiling 64
2.3.9 Experimental evolution 64
2.4 Methodology specific to Chapter 5: Southeast Asian *Shigella* from a historical perspective 66
2.4.1 Sample collection and genome sequencing 66
2.4.2 *Shigella flexneri* molecular serotyping scheme 69
2.4.3 Phylogenetic reconstructions 71

2.4.4 Bayesian phylogenetic inferences 72
2.4.5 Inter-regional dissemination of *Shigella* lineages 74
2.5 Methodology specific to Chapter 6: Assessing the gut microbiota perturbations during the early phase of infectious diarrhoea in Vietnamese children 75
2.5.1 Study design, sample collection and microbiological procedures 75
2.5.2 DNA extraction and sequencing 76
2.5.3 EMIRGE assembly of 16S rRNA amplicons and OTU clustering 77
2.5.4 Data analysis 79
2.5.5 Clustering into Community State Type (CST) 80
2.5.6 Statistical analysis and regression modelling 81
2.5.7 Evaluating differential abundance 82
2.5.8 Interaction network construction 83
3. The introduction and establishment of fluoroquinolone resistant *Shigella sonnei* into Bhutan 84
3.1 Abstract 84
3.2 Introduction 85
3.3 Results 87
3.3.1 Bhutanese *Shigella sonnei* lay within lineage III (Central Asia clade) 87
3.3.2 Temporal phylogenetic reconstruction of *Shigella sonnei* in Bhutan 91
3.3.3 Mutations specific to Bhutanese *Shigella sonnei* 94
3.3.4 The accessory genome in Bhutanese *Shigella sonnei* 96
3.4 Discussion 100

4. The emergence and global spread of fluoroquinolone resistant *Shigella sonnei*.....103

4.1 Abstract.....103

4.2 Introduction.....104

4.3 Results.....106

4.3.1 Fluoroquinolone resistant *Shigella sonnei* in global context.....106

4.3.2 The emergence of the fluoroquinolone resistant *Shigella sonnei* clone.....108

4.3.3 Phylogeography and the origin of fluoroquinolone resistant *Shigella sonnei* 119

4.3.4 Accessory genome investigation.....121

4.3.5 Reductive evolution in CenAsiaIII *Shigella sonnei*.....124

4.3.6 Substitution mutation analysis.....130

4.3.7 *Shigella sonnei* experimental evolution.....132

4.4 Discussion.....137

5. Southeast Asian *Shigella* from a historical perspective.....142

5.1 Abstract.....142

5.2 Introduction.....143

5.3 Results.....144

5.3.1 Southeast Asian *Shigella* in global context.....144

5.3.2 Co-circulations of multiple *Shigella* lineages in Southeast Asia.....150

5.3.3 Antimicrobial resistance pattern underlies *Shigella* clonal replacement in Southeast Asia.....171

5.4 Discussion.....183

6. Assessing gut microbiota perturbations during the early phase of infectious diarrhoea in Vietnamese children 188

6.1 Abstract 188

6.2 Introduction 189

6.3 Results 191

6.3.1 16S rRNA gene sequencing of faecal diarrhoeal samples 191

6.3.2 Microbial structures within the faecal samples 193

6.3.3 Factors affecting microbial structure in the faeces of children with diarrhoea 204

6.3.4 Changes in taxonomic abundance in the faeces of children with diarrhoea 207

6.3.5 Taxonomic changes associated with bacterial induced diarrhoea and dysentery 214

6.4 Discussion 215

7. General Discussion 221

7.1 The surge of fluoroquinolone resistant *Shigella sonnei* 221

7.2 Historical epidemiology of *Shigella* in Southeast Asia 224

7.3 Microbiota perturbation under infectious diarrhoea 225

8. References 227

9. Appendices 280

9.1 Publications as first author during DPhil candidature 280

List of Tables

Table 2.1: Summary of isolates used in Southeast Asia *Shigella* evolution analysis 68

Table 2.2: Geno-serotyping scheme for *Shigella flexneri*.....70

Table 2.3: Input alignment for each *Shigella* lineage used for Bayesian phylogenetic inferences.....73

Table 2.4: Input information for stochastic mapping of geographical states, separately for each major *Shigella* lineage.....74

Table 3.1: The nonsynonymous mutations separating the MRCA of all Bhutanese *Shigella sonnei* from other CenAsiaIII clade strains.....95

Table 4.1: Summary of sequences used to investigate the emergence of fluoroquinolone resistant *Shigella sonnei* 109

Table 4.2: List of lineage-defining pseudogenes and gene loss in 395 CenAsiaIII *Shigella sonnei*.....126

Table 4.3: Genes likely under natural selection in CenAsiaIII *Shigella sonnei* population 131

Table 4.4: Results on antimicrobial susceptibility and mutation analysis of the *Shigella sonnei* experimental evolution.....135

Table 5.1: Summary of findings from Bayesian phylogenetic inferences of major *Shigella* lineages 157

Table 6.1: Summary of aetiologies associated with 142 diarrhoeal samples contributing to this study.....192

Table 6.2: Demographic and clinical characteristics of 136 diarrhoeal individuals included in the multinomial regression analysis, stratified by the membership of bacterial community state type (CST).....205

Table 6.3: Results of multinomial logistic regression for association of demographic and clinical predictors for classification of community state types (CSTs) 206

List of Figures

Figure 1.1: Progression of gut microbiota in infancy 4

Figure 1.2: Proposed model of gut microbiota succession after diarrhoea 8

Figure 1.3: The multi-step process of *Shigella*'s molecular pathogenesis 14

Figure 1.4: Epidemiology of *Shigella* species 18

Figure 1.5: The phylogenetic structure of four *Shigella* species 24

Figure 1.6: Phylogenomics of *Shigella* species 27

Figure 1.7: A comparative gene map of the *Shigella* virulence plasmid 30

Figure 1.8: Global population structure of *Shigella sonnei* 36

Figure 1.9: Circulation of *Shigella sonnei* in Vietnam 38

Figure 1.10: Schematic representation of *Shigella flexneri*'s serotype conversion pathway 41

Figure 1.11: The global phylogenetic structure of *Shigella flexneri* 45

Figure 3.1: The phylogenetic structure of Bhutanese *Shigella sonnei* in the context of the global phylogeny and CenAsiaIII clade 88

Figure 3.2: Recombination features within the genomes of CenAsiaIII *Shigella sonnei* 90

Figure 3.3: Temporal phylogenetic reconstruction of Bhutanese *Shigella sonnei* between 2011 and 2013 92

Figure 3.4: Novel recombinant hybrid colicin biosynthesis cluster encoded on pSSE3 98

Figure 4.1: The phylogenetic structure of ciprofloxacin resistant *Shigella sonnei* in an international context 107

Figure 4.2: The phylogenetic structure of CenAsiaIII *Shigella sonnei* 112

Figure 4.3: Plasmid profile and antimicrobial resistance genes among 395 CenAsiaIII *Shigella sonnei* 114

Figure 4.4: Recombination features within the genomes of 395 CenAsiaIII *Shigella sonnei* isolates 115

Figure 4.5: Temporal phylogenetic reconstruction of CenAsiaIII *Shigella sonnei* overlaid with pseudogenes 116

Figure 4.6: Bayesian skyline plot showing the change in population size of CenAsiaIII *Shigella sonnei* 118

Figure 4.7: Geographical spreads of CenAsiaIII *Shigella sonnei* 120

Figure 4.8: Plasmid profile and colicin families present in 395 CenAsiaIII *Shigella sonnei* 123

Figure 4.9: Cytokine measurements from plasma of patients infected with *Shigella sonnei* 128

Figure 5.1: Overview of Southeast Asian *Shigella* sequences used in this study 145

Figure 5.2: The global phylogenetic structure of *Shigella flexneri* 148

Figure 5.3: The global phylogenetic structure of *Shigella sonnei* 149

Figure 5.4: Summary of *Shigella* sampling timeframe employed in this study 151

Figure 5.5: Temporal phylogenetic signal for *Shigella flexneri* 152

Figure 5.6: Temporal phylogenetic signal for *Shigella sonnei* 153

Figure 5.7: Recombination features within representative Southeast Asia *Shigella flexneri* lineages 155

Figure 5.8: Recombination features within representative Southeast Asia *Shigella sonnei* lineages 156

Figure 5.9: Circulations of different *Shigella* lineages in Southeast Asia..... 158

Figure 5.10: Migration network of *Shigella* within Southeast Asia..... 159

Figure 5.11: Temporal phylogenetic reconstructions of *S. flexneri* Lineage 3 161

Figure 5.12: Temporal phylogenetic reconstructions of *S. sonnei* Lineage 3..... 163

Figure 5.13: Temporal phylogenetic reconstructions of *S. flexneri* Lineage 1 and 3Xv 165

Figure 5.14: Temporal phylogenetic reconstructions of *S. flexneri* Lineage 2 166

Figure 5.15: Temporal phylogenetic reconstructions of *S. sonnei* Lineage 1 and Lineage 3 clade early-SEA 169

Figure 5.16: Temporal phylogenetic reconstructions of *S. flexneri* Lineage 7 and 8 .. 170

Figure 5.17: Trend of antimicrobial resistance genotype in 990 Southeast Asian *Shigella flexneri* 174

Figure 5.18: Clonal replacement of *Shigella flexneri* in Southeast Asia..... 176

Figure 5.19: Trend of antimicrobial resistance genotype in 880 Southeast Asian *Shigella sonnei*..... 180

Figure 5.20: Clonal replacement of *Shigella sonnei* in Southeast Asia 182

Figure 6.1: Estimating the number of clusters for the microbial compositions of 199 faecal samples 195

Figure 6.2: Dissimilarity in bacterial compositions of four Community State Types (CSTs)..... 196

Figure 6.3: The impact of diarrhoea and demographics on the gut bacterial configurations 197

Figure 6.4: Examination of CST clustering with Random Forest classification 198

Figure 6.5: Genera defining the clustering of CSTs 199

Figure 6.6: The bacterial compositions of the four CSTs 200

Figure 6.7: Shannon diversity index for 195 faecal samples, classified by diarrhoeal status and CST membership 203

Figure 6.8: Bacterial taxa showing significantly different abundance among the examined classes 209

Figure 6.9: Bacterial taxa with decreased abundance during the early phase of diarrhoea 211

Figure 6.10: Correlation network of the healthy and diarrhoeal microbiome 213

Abbreviation

ACC	Acetyl-CoA carboxylase
AFRIMS	Armed Forces Research Institute of Medical Sciences
AIC	Akaike information criterion
AMR	Antimicrobial resistance
ANOVA	Analysis of variance
ARD	Asymmetric model of character change
asw	Average silhouette width
BEAST	Bayesian Evolutionary Analysis of Sampling Trees
BLAST	Basic Local Alignment Search Tool
BWA	Burrows-Wheeler Aligner
CAP	Constrained analysis of principal coordinates
CDI	*Clostridium difficile* infection
CenAsiaIII	Central Asia III
CH1	Children Hospital 1
CH2	Children Hospital 2
CIP	Ciprofloxacin
cipR	Ciprofloxacin resistant
cipS	Ciprofloxacin sensitive
COMRU	Cambodia-Oxford Medical Research Unit
CST	Community state type
DALY	Disability-adjusted life year
dN/dS	Nonsynonymous to synonymous mutations

DNA	Deoxyribose nucleic acid
EAEC	Enteroaggregative *Escherichia coli*
EHEC	Enterohemorrhagic *Escherichia coli*
EIEC	Enteroinvasive *Escherichia coli*
EMIRGE	Expectation Maximization Iterative Reconstruction of Genes from the Environment
ER	Endoplasmic reticulum
ESBL	Extended-spectrum beta-lactamase
ESS	Effective sample size
FFP	Feature frequency profile
FMT	Faecal microbiota transplantation
FQ	Fluoroquinolone
GATK	Genome Analysis Tool Kit
GBD	Global Burden of Disease
GEMS	Global Enteric Multicenter Study
GERMS	GIS Efficient Rapid Microbial Sequencing
GIS	Genome Institute of Singapore
GTR	General Time Reversible
gtr	glucosyltransferase
HCMC	Ho Chi Minh city
HGT	Horizontal gene transfer
hierBAPS	Hierarchical Bayesian clustering
HMP	Human Microbiome Project

HOMD	Human Oral Microbiome Database
HPD	Highest posterior density
HTD	Hospital for Tropical Diseases
HUS	Haemolytic uremic syndrome
IEC	Intestinal epithelial cell
IIOD	Induced iso-osmotic diarrhoea
IQR	Interquartile range
IS	Insertion sequence
JDWNRH	Jigme Dorji Wangchuk National Referral Hospital
K3P	Kimura 3 parameters
LDC	Lysine decarboxylase
LOMWRU	Lao-Oxford-Mahosot Hospital-Wellcome Trust Research Unit
LRT	Likelihood ratio test
LSU	Large subunit
MCC	Maximum clade credibility
MCMC	Markov Chain Monte Carlo
MDR	Multidrug resistant
MIC	Minimum inhibitory concentration
MLST	Multi-locus sequence typing
MRCA	Most recent common ancestor
MRE	extend majority rule
MSM	Men who have sex with men
NAD	Nicotinamide adenine dinucleotide

NADPH	Nicotinamide adenine dinucleotide phosphate
NAL	Nalidixic acid
oac	O-acetylation
OAS	O-acetyl-L-serine
OOB	Out-of-bag
OTU	Operational taxa unit
OUCRU	Oxford University Clinical Research Unit
OxTREC	Oxford Tropical Research Ethics Committee
PAI	Pathogenicity Island
pam	Partitioning around meloids
PAMP	Pathogen-associated molecular pattern
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
PEtN	phosphoethanolamine
PFGE	Pulse field gel electrophoresis
QRDR	Quinolone resistance determining region
RDP	Ribosomal Database Project
Rha	Rhamnose
ROS	Reactive oxygen species
rRNA	ribosomal ribonucleic acid
SCFA	Short chain fatty acid
Sd1	*Shigella dysenteriae* serotype 1
SEA	Southeast Asia

SPATE	Serine protease autotransporter of Enterobacteriaceae
SRL	*Shigella* resistance locus
SSU	Small subunit
STEC	Shiga-toxin producing Escherichia coli
STI	Sexually-transmitted infection
Stx	Shiga toxin
SXT	Co-trimoxazole (Trimethoprim/sulfamethoxazole)
T3SS	Type 3 secretion system
TLR5	Toll-like receptor 5
TVM	Transversion model
UK	United Kingdom
USA	United States of America
WAZ	Weight-for-age Z score
WGS	Whole genome sequencing
WHO	World Health Organization
WTSI	Wellcome Trust Sanger Institute

Publications during DPhil candidature

1. **Chung The, H.,** Karkey, A., Pham Thanh, D., Boinett, C. J., Cain, A. K., Ellington, M., Baker, K. S., Dongol, S., Thompson, C., Harris, S. R., Jombart, T., Le Thi Phuong, T., Tran Do Hoang, N., Thanh, T. H., Shretha, S., Joshi, S., Basnyat, B., Thwaites, G., Thomson, N. R., Rabaa, M. A. and Baker, S. (2015) 'A high-resolution genomic analysis of multidrug- resistant hospital outbreaks of *Klebsiella pneumoniae*', EMBO Molecular Medicine, 7(3), pp. 227–239.
2. **Chung The, H.**, Pham Thanh, D., Holt, K. E., Thomson, N. R. and Baker, S. (2016) 'The genomic signatures of *Shigella* evolution, adaptation and geographical spread', Nature Reviews Microbiology. 14(4), pp. 235–250.
3. **Chung The, H.**, Rabaa, M. A., Pham Thanh, D., Lappe, N. De, Cormican, M., Valcanis, M., Howden, B. P., Wangchuk, S., Bodhidatta, L., Mason, C. J., Nguyen, T. N. T., Thuy, D. V., Thompson, C. N., Lan, N. P. H., Vinh, P. V., Thanh, T. H., Turner, P., Sar, P., Thwaites, G., Thomson, N. R., Holt, K. E. and Baker, S. (2016) 'South Asia as a reservoir for the global spread of ciprofloxacin resistant *Shigella sonnei*', PLoS Medicine, 13(8), pp. 1–12..
4. **Chung The, H.,** Rabaa, M., Pham Thanh, D., Ruekit, S., Wangchuk, S., Dorji, T., Tshering, K. P., Nguyen, T. N. T., Vinh, P., Thanh, T., Minh, C., Turner, P., Sar, P., Thwaites, G., Holt, K. E., Thomson, N. R., Bodhidatta, L., Mason, C. J. and Baker, S. (2015) 'Introduction and establishment of fluoroquinolone resistant *Shigella sonnei* into Bhutan', Microbial Genomics, 1(6), pp. 1–11.
5. **Chung The, H.**, Sessions, P. F. de, Jie, S., Thanh, D. P., Thompson, C. N., Minh, C. N. N., Chu, C. W., Tran, T.-A., Thomson, N. R., Thwaites, G. E., Rabaa, M. A.,

Hibberd, M. and Baker, S. (2017) 'Assessing gut microbiota perturbations during the early phase of infectious diarrhea in Vietnamese children', Gut Microbes. 0(0), pp. 1–17.

6. **Chung The, H.**, and Baker, S. (2018). 'Out of Asia: The independent rise and global spread of fluoroquinolone resistant *Shigella'*, Microbial Genomics, 4, pp. 1-6.
7. Darton, T. C., Tuyen, H. T., **Chung The, H**., Newton, P. N., Phetsouvanh, R., Davong, V., Campbell, J. I., Hoang, N. V. M., Thwaites, G. E., Parry, C. M., Thanh, D. P., and Baker, S. (2018) 'Azithromycin resistance in *Shigella* spp. in Southeast Asia', Antimicrobial Agents and Chemotherapy, (January).
8. Leung, J. M., Budischak, S. A., **Chung The, H.**, Hansen, C., Bowcutt, R., Neill, R., Shellman, M., Loke, P. and Graham, A. L. (2018) 'Rapid environmental effects on gut nematode susceptibility in rewilded mice', PLoS Biology, pp. 1–28.
9. Hong Chau, T. T., Chau, N. N. M., Hoang Le, N. T., **Chung The, H.**, Voong Vinh, P., To, N. T. N., Ngoc, N. M., Tuan, H. M., Ngoc, T. L. C., Kolader, M.-E., Farrar, J. J., Wolbers, M., Thwaites, G. E., Baker, S. (2017) A Double Blind, Randomized, Placebo-Controlled Trial of *Lactobacillus acidophilus* for the Treatment of Acute Watery Diarrhea in Vietnamese Children. The Pediatric Infectious Disease Journal.
10. Pham Thanh, D., Thanh Tuyen, H., Nguyen Thi Nguyen, T., **Chung The, H.,** Wick, R. R., Thwaites, G. E., Baker, S. and Holt, K. E. (2016) 'Inducible colistin resistance via a disrupted plasmid-borne mcr-1 gene in a 2008 Vietnamese *Shigella sonnei* isolate', Journal of Antimicrobial Chemotherapy, 71(8), pp. 2314–2317.
11. Thompson, C. N., Vu Thieu, N. T., Vinh, P. V., Duc, A. N., Wolbers, M., Vinh, H., Campbell, J. I., Ngoc, D. T. T., Hoang, N. V. M., Ha Thanh, T., **Chung The, H**.,

Nguyen Thi Nguyen, T., Nguyen Phu Lan, H., Parry, C. M., Chau, N. V. V., Thwaites, G., Thanh, D. P. and Baker, S. (2016) 'Clinical implications of reduced susceptibility to fluoroquinolones in paediatric *Shigella sonnei* and *Shigella flexneri* infections', Journal of Antimicrobial Chemotherapy, 71(3), pp. 807–815.

1 Introduction

A major part of this section pertaining the pathogenesis, genomic, evolution, and epidemiology of *Shigella* species has been published during my DPhil in a review.

Chung The, H., Pham Thanh, D., Holt, K. E., Thomson, N. R. and Baker, S. (2016) 'The genomic signatures of *Shigella* evolution, adaptation and geographical spread', *Nature Reviews Microbiology*. Nature Publishing Group, 14(4), pp. 235–250.

1.1 Global burden of diarrhoeal diseases

Estimates from the Global Burden of Disease (GBD) Study concluded that diarrhoea was a major cause of mortality among all age groups, leading to more than 1.3 million deaths globally in 2015 (Troeger *et al.*, 2017). The burden of diarrhoea is highest among children aged under five years, causing approximately 500,000 deaths annually and ranking it as the fourth most common cause of mortality in young children worldwide (WHO, 2016). Developing and low-middle income regions, such as sub-Saharan Africa and South Asia, are disproportionally affected by diarrhoeal diseases due to inadequate access to clean water, sanitation, and healthcare facilities. While diarrhoea related mortality decreased substantially by 39% over the past decade, the incidence of diarrhoea decreased more slowly by 10.4% over the same period, culminating in ~950 million diarrhoea episodes in young children (< 5 years) annually (Troeger *et al.*, 2017) As a result, diarrhoea remains a substantial cause of disability in this age group, with estimated disability adjusted life years (DALYs) of greater than 45 million (WHO, 2016). Repeated diarrhoeal episodes in children cumulatively increase the risk of malnutrition and

stunting, which are associated with cognitive impairment and development of cardiovascular diseases and glucose intolerance in adulthood (Checkley *et al.*, 2008; Deboer *et al.*, 2012; Guerrant *et al.*, 2013). Therefore, aside from causing tremendous morbidity in young children, diarrhoeal diseases can greatly exhaust societal resources, especially in impoverished regions, making it a pressing public health issue worldwide.

1.2 The diarrhoeal gut microbiota

1.2.1 The human gut microbiota and its development during infancy

The human body is populated with immensely rich and diverse microbial communities, with the large intestine harbouring the greatest density of bacteria (The Human Microbiome Project Consortium, 2012). These microbial communities, collectively termed the microbiota or microbiome, are increasingly recognized as an indispensable component of the human body. Harbouring thousands of bacterial species, the human gut microbiota carries out a myriad of important functions, including nutrient metabolism (particularly of indigestible carbohydrates), xenobiotic metabolism, vitamin production, protection against pathogen colonization and infection, and modulation of the immune response (Sekirov, Russell and Antunes, 2010; Jandhyala *et al.*, 2015). Despite the overwhelming diversity of the prokaryotic world, the majority of bacteria inhabiting the large intestine in healthy adults belong to two phyla, Firmicutes and Bacteroidetes. Most of these are obligate anaerobes, and they are rarely found proliferating in habitats other than the intestinal environment, thus suggesting that members of the gut microbiota have evolved to adapt to this specific niche (Browne *et al.*, 2017).

The dynamics of human gut microbiota are most apparent during its development in infancy, with its establishment starting at birth (Yatsunenko *et al.*, 2012). However, this initial bacterial colonization is largely dependent on the mode of delivery. The intestine of vaginally delivered neonates are first colonized by Enterobacteriaceae (mostly *Escherichia*), but the microbiota is rapidly enriched with *Bifidobacterium* and *Bacteroides* within a few days (Fig 1.1) (Bäckhed *et al.*, 2015; Dogra *et al.*, 2015). High resolution organism profiling indicates that these taxa likely originate from the mother's gut microbiota (Nayfach *et al.*, 2016). Sustainment of *Bifidobacterium* is further promoted in breastfed infants since human breast milk contains indigestible oligosaccharides, which are preferentially imported and then degraded by members of this genus (Charbonneau *et al.*, 2016). This strategy ensures such nutrients from breast milk are inaccessible to other pathogens and commensals. Later, the cessation of breastfeeding proves to be key in transitioning the gut microbiota into the adult-like configuration, which is dominated by Firmicutes and Bacteroidetes capable of metabolizing a diverse range of nutrient intake (Fig 1.1) (Bäckhed *et al.*, 2015). While the class Bacteroidia remains genetically similar to that inherited from the mother as the gut matures, members of the class Clostridia display more genetic heterogeneity (Nayfach *et al.*, 2016). This emphasizes the importance of environmental transmission in shaping the microbiota in maturation (Browne *et al.*, 2017).

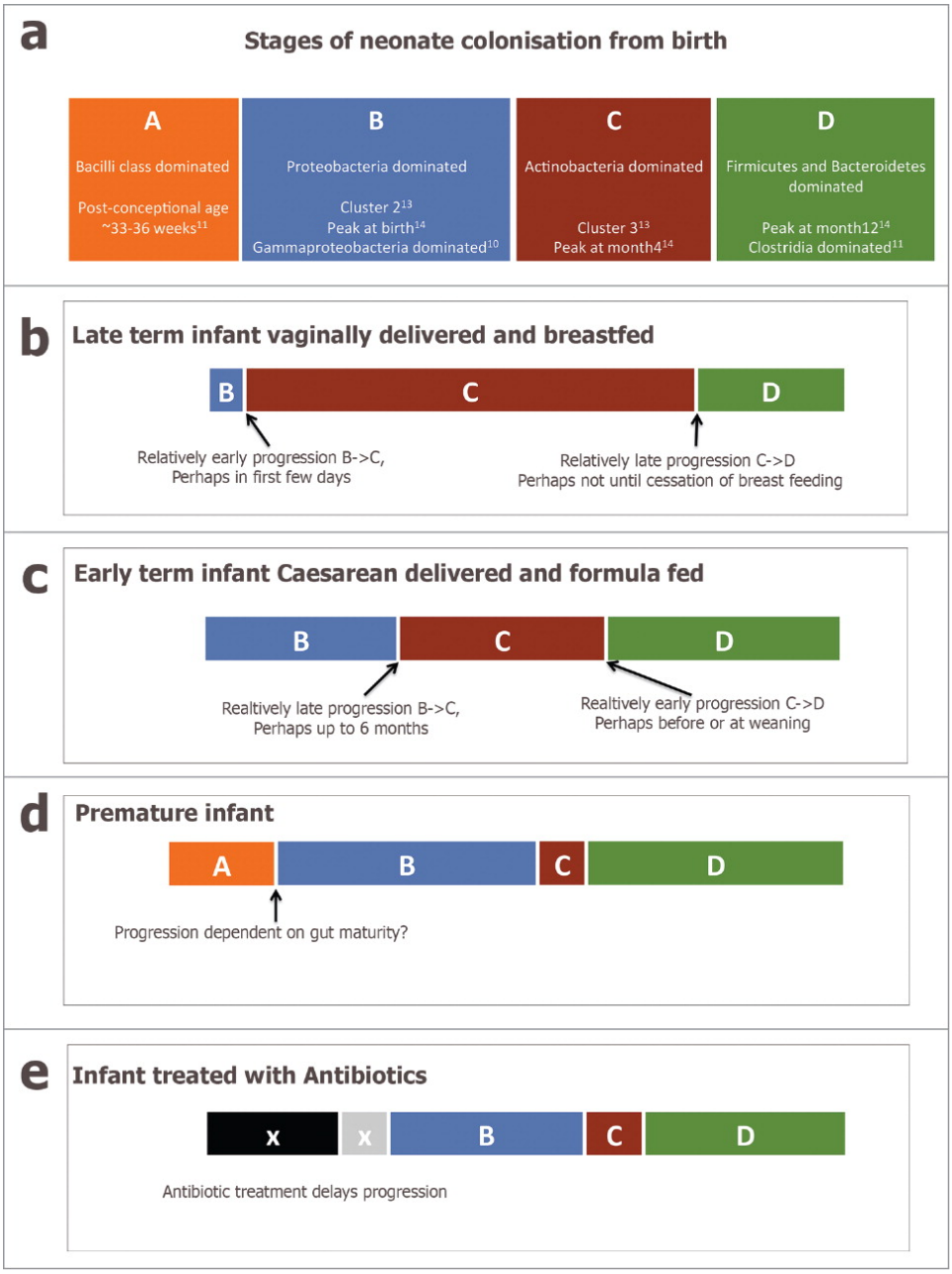


Figure 1.1 Progression of gut microbiota in infancy (Adopted from (Dogra *et al.*, 2015)). (a) Simplified representation of the infant faecal microbiota development during the first year of life, including stages A – D. (b-e). Proposed modifications to the progression due to intrinsic and extrinsic factors.

In contrast, the first gut colonizers of infants born by Caesarean section (C-section) are predominant with taxa found on the skin, mouth, and the surrounding environment, such as *Enterobacter*, *Staphylococcus*, and *Streptococcus.* Vertical transmission of the gut microbiota from mothers to infants is observed less frequently in C-section neonates, particularly for the genus *Bacteroides*. This may explain the absence or delayed colonization of Bacteroidetes in C-section infants, in comparison with ones delivered vaginally in the same age group (Jakobsson *et al.*, 2014; Bäckhed *et al.*, 2015). Infants receiving formula feeding harbour a microbiota characterized as 'more mature' than their exclusively breastfed counterparts, prompting an earlier transition into the adult-like composition (Fig 1.1) (Bäckhed *et al.*, 2015; Dogra *et al.*, 2015). In children suffering from severe acute malnutrition, the gut microbiota remains in an immature state due to the decreased abundance of key taxa such as *Faecalibacterium* and *Ruminococcus* (Subramanian *et al.*, 2014). These findings demonstrate that the intestinal microbial communities are highly variable in response to intrinsic and environmental factors. Upon reaching the adult-like composition at around three years of age, the gut microbiota is more stable and resilient to external impacts. Nevertheless, it is still subjected to alternations exerted by diet, gastrointestinal infections, and use of antibiotics (David, Materna, *et al.*, 2014; David, Maurice, *et al.*, 2014). This makes the definition of a standard core 'healthy' microbiota implausible.

1.2.2 Impact of diarrhoea on gut microbiota

The versatile functions of the gut microbiota offer new perspectives to understand various aspects essential to human health. Numerous studies have attempted to determine specific

microbial signatures in association with a variety of clinical conditions (Gilbert *et al.*, 2016). These initially focused on gastrointestinal syndromes (such as inflammatory bowel disease (Halfvarson *et al.*, 2017)) and metabolic diseases (obesity (Turnbaugh *et al.*, 2006) and type-2 diabetes (Pedersen *et al.*, 2016)), but later expanded into conditions such as cancer (Kostic *et al.*, 2012), major depression (Naseribafrouei *et al.*, 2014), and autism spectrum disorder (Vuong and Hsiao, 2017). In high income settings, growing concerns over the incidence of *Clostridium difficile* infections (CDI), frequently as a consequence of an extended antimicrobial usage, have instigated the use of microbiome-oriented treatments. Specifically, faecal microbiota transplant (FMT) is now recommended for CDI patients, showing a diarrhoea resolution rate surpassing 80% (Drekonja *et al.*, 2015). A multidisciplinary and comprehensive study has recently identified *Clostridium scindens* as a candidate mediating CDI resistance in humans, and its role was later confirmed in animal model, prompting the possibility of evidence based microbiome intervention (Buffie *et al.*, 2014).

Despite the advancements made in these fields, microbial dysbiosis remains insufficiently characterized for infectious diarrhoea in high incidence settings. The aforementioned highly dynamic succession of microbial colonization in young children also confounds analysis in this target group (Bäckhed *et al.*, 2015; Dogra *et al.*, 2015). Nevertheless, several 16S rRNA gene profiling studies have shown consistent patterns in the initial gut microbiota response following acute diarrhoeal episodes. Enterobacteriaceae and/or *Streptococcus* are found in greater abundance in the faecal microbiome of diarrhoeal children, and this is coupled with a reduction in specific Firmicutes and Bacteroidetes

colonizers (Monira *et al.*, 2013; Pop *et al.*, 2014). In comparison to healthy controls, the diarrhoeal gut microbiome displays a uniform drop in diversity and a shift to facultative anaerobes. A pioneering shotgun metagenomics study on children infected with *Vibrio cholerae* and enterotoxigenic *Escherichia coli* (ETEC) concluded that following diarrhoea, the gut microbiota undergoes an orderly succession (David *et al.*, 2015). Initial predominance of the bacterial pathogen is quickly replaced by the transient colonization of facultative anaerobes (*Escherichia*, *Enterococcus*, and *Streptococcus*), even one day after the onset of symptoms. This is followed by the enrichment of strict anaerobes (*Bacteroides*, *Prevotella*, *Roseburia*) in the recovery phase, and returning the gut microbiota to a configuration similar to that found in non-diarrhoeal individuals one month post diarrhoea (Fig 1.2). Separate investigations of the diarrhoeal gut microbiome showed increases in the abundance of *Bacteroides* and Firmicutes taxa in recovery phase, emphasizing that this phenomenon may be universal (Hsiao *et al.*, 2014; Becker-dreps *et al.*, 2015). Detailed inspection of the differential gene content points to the immediate increase of low-affinity oxidases in the early phase of diarrhoea, indicating a switch to a more oxygenated gut environment, thus favouring the growth of facultative anaerobes (David *et al.*, 2015). As oxygen is gradually consumed by aerobic respiration, strict anaerobes begin to thrive and compete for the nutrient resource. This process is accelerated by the vast repertoire of carbohydrate metabolism genes found in Bacteroidetes and Firmicutes.

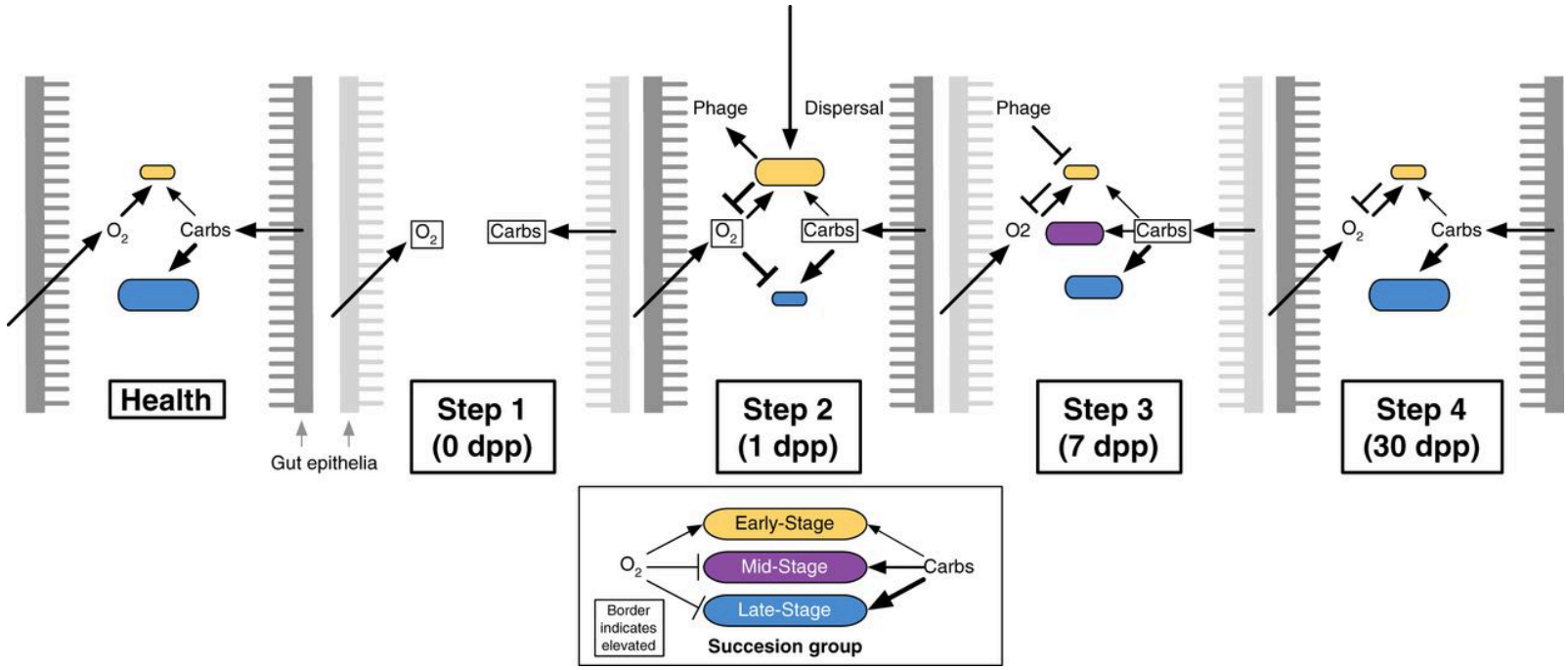


Figure 1.2 Proposed model of gut microbiota succession after diarrhoea (Adopted from (David *et al.*, 2015)). Three groups of bacteria are shown to colonize the gut: Early-Stage (*Escherichia*, *Enterococcus*, *Streptococcus*), Mid-Stage (*Bacteroides*), and Late-Stage (*Prevotella*, *Roseburia*) in an orderly manner. Diarrhoea diminishes the gut bacterial communities, leading to the accumulation of host secreted oxygen and carbohydrate (Step 1). Early-phase bacteria tolerate the increasingly oxygenated environment and colonize the gut (Step 2). Aerobic respiration gradually depletes the oxygen level, favouring the proliferation of Mid and Late-Stage bacteria (Step 3). More efficient utilization of carbohydrate resource by Late-Stage bacteria returns the gut microbiota to the 'Health' Stage (Step 4).

1.3 The bacterial genus *Shigella*

The recent Global Enteric Multicenter Study (GEMS), an expansive case-control study of moderate to severe paediatric diarrhoeal disease, identified ETEC and *Shigella* as the two most common bacterial pathogens in sub-Saharan Africa and South Asia (Kotloff *et al.*, 2012, 2013), with *Shigella* spp. being the most prevalent pathogen detected among children between 24 and 59 months old (~1 incidence per 100 child-years) (Kotloff *et al.*, 2013). Historical data suggest that there were ~165 million cases of *Shigella* disease annually mainly in low-income countries and in children younger than 5 years between 1966 and 1997, resulting in 1.1 million deaths per year worldwide (Kotloff *et al.*, 1999). During 1990-2009, it was estimated that *Shigella* spp. cause ~125 million disease cases annually (Bardhan *et al.*, 2010). The incidence of shigellosis was approximated to be 13.2/1,000/year in children aged under five years in Asia (von Seidlein *et al.*, 2006). Globally, deaths attributed to shigellosis reach approximately 160,000 per annum, and a third of this burden occurs in young children (Troeger *et al.*, 2017). Importantly, despite the continued high incidence of *Shigella* infections, the mortality rate per case has dropped significantly; this is likely due to the disappearance of epidemics associated with the highly pathogenic *S. dysenteriae* and improved, more rapid treatment (Bardhan *et al.*, 2010).

1.3.1 Nomenclature and classification

The highly pathogenic and exotoxin producing *Shigella dysenteriae* was first described in 1897 by Kiyoshi Shiga, who isolated this species from the stool sample of a patient with epidemic dysentery in Japan (Shiga, 1898; Trofa *et al.*, 1999). The genus was expanded

soon after: *Shigella flexneri* was identified in 1899, *Shigella sonnei* in 1906, and *Shigella boydii* in 1921 (Flexner, 1900; Barceloux, 2008). Current serological classification divides the genus *Shigella* into four species (also known as subgroups), which are further subdivided into serotypes by type-specific antigens: *S. dysenteriae* (subgroup A) has 15 serotypes; *S. flexneri* (subgroup B) has 19 serotypes and sub serotypes; *S. boydii* (subgroup C) has 20 serotypes; and *S. sonnei* (subgroup D) consists of a single serotype.

1.3.2 Laboratory identification of *Shigella*

Shigella is a facultatively anaerobic Gram-negative bacterium, belonging to the Enterobacteriaceae family. It is rod-shaped, non-motile, non-sporulating and non- or late-lactose fermenting, and classical taxonomy places all the *Shigella* species into one major group, which is distantly related to *E. coli* (Dodd and Jones, 1982). However, even on their initial characterization, the biochemical and morphological proximity of members of the genus *Shigella* to *E. coli* was noted (Shiga, 1898). Moreover, biochemical traits are varying in both of these genera. For example, indole production is present in most *E. coli* but only absent for specific *Shigella* serotypes, including *S. sonnei*, *S. dysenteriae* 1 and *S. flexneri* 6 (Rezwan, Lan and Reeves, 2004). *S. dysenteriae* does not ferment mannitol while most members of the remaining *Shigella* species do (Public Health England, 2015). These make differentiation between the two genera a daunting task, but a few biochemical and physiological differences are recommended to achieve this. All *Shigella* spp. are negative for lysine decarboxylation (LDC) and are nonmotile, while most *E. coli* biovars are positive for these two traits (Falkow, 1958; van den Beld and Reubsaet, 2012). However, enteroinvasive *E. coli* (EIEC) is known to be indistinguishable from

Shigella in these two tests due to their similar patho-adaptation, and there is currently no known phenotypic assay for confident differentiation (van den Beld and Reubsaet, 2012).

In addition, *Shigella*-specific molecular markers are commonly used in Polymerase Chain Reaction (PCR) for optimal detection from clinical and environmental samples. The commonly used targets comprise of virulence genes found on the *Shigella* large virulence plasmid (invasion plasmid locus, *ipl*) or the multi-copy chromosomal *ipaH* (Sethabutr *et al.*, 1993; Niyogi, 2005). Indeed, *ipaH* PCR has exhibited high sensitivity and helped researchers to accurately evaluate the burden of shigellosis locally and globally, compared to the usually underestimated figure produced by the traditional culture approach (Vu *et al.*, 2004; Lindsay *et al.*, 2013). However, the specificity of these reactions still extends to the EIEC, so they could not be used as a definitive diagnostic for *Shigella* only. Another valuable approach is serotyping, which depends on the agglutinations of *Shigella*-specific O-antigens and commercial antisera. Since EIEC possesses only a range of *E. coli* O-serotypes, which do not show cross-reactivity with *Shigella* antisera, serotyping is a powerful tool to separate *Shigella* from EIEC as well as to detail the organism's serotype and subtype (van den Beld and Reubsaet, 2012).

1.3.3 Molecular mechanism of *Shigella* pathogenesis

Infection with *Shigella* spp. usually results in self-limiting diarrhoea, initially watery and then bloody and/or mucus containing. Pathogenesis is tightly orchestrated by the virulence factors encoded on the virulence plasmid, which can be as large as ~220 kbp (Buchrieser *et al.*, 2000). It contains the conserved 30kbp *mxi-spa* locus, which encodes

the Mxi-Spa type 3 secretion system (T3SS), and a plethora of effector proteins, including the invasion plasmid antigens (IpaABCDH). The secretion apparatus allows for a complex interaction between the organism and the host cell, ultimately resulting in a disruption of the intestinal mucosa and the distinctive symptoms of bacillary dysentery. This has been proposed to be a multistep process (Fig 1.3) (Schroeder and Hilbi, 2008; Carayol and Tran Van Nhieu, 2013). First, *Shigella* crosses the intestinal epithelium via M cells (Sansonetti *et al.*, 1996). This process, termed transcytosis, allows *Shigella* to cross the host's physical protective barrier and exposes it to macrophages in the submucosa of the gastrointestinal tract. *Shigella* is phagocytized, but escapes destruction by inducing rapid macrophage apoptosis through a caspase-1 dependent pathway (Zychlinsky, Prevost and Sansonetti, 1992; Hilbi *et al.*, 1998). After its release, the bacterium comes into contact with the basolateral side of intestinal epithelial cells (IECs) and triggers effector-mediated endocytosis. The Mxi-Spa T3SS needle-like structure is coated by the protein IpaB at the tip, which has a high affinity to cholesterol-rich areas of eukaryotic cell membranes (Lafont *et al.*, 2002). *Shigella*'s adhesion to IECs is achieved by binding of IpaB and IpaBCD to the host hyaluronan receptor CD44 and $\alpha_5\beta_1$ integrin, respectively (Watarai, Funato and Sasakawa, 1996; Skoudy *et al.*, 2000). IcsA has been recently shown to function as an indispensable adhesin to promote contact with host IECs (Brotcke Zumsteg *et al.*, 2014). This close contact facilitates the fusion of the T3SS apparatus and the host cell membrane, leading to the translocation of further effectors. Successful engulfment requires extensive reorganization of the host cell cytoskeleton and modulation of other intracellular pathways, involving the activity of the effectors IpaC, IpgB1, IpgD, IpaA and VirA in concert (Niebuhr, Giuriato and Pedron, 2002; Yoshida

and Katayama, 2002; Handa *et al.*, 2007; Mounier *et al.*, 2009; Park *et al.*, 2011). After invasion, *Shigella* quickly lyses its surrounding phagosome by activation of IpaC, IpaD, IpaB and IpaH, of which IpaC has been shown to play the pivotal role (Fernandez-prada *et al.*, 2000; Harrington, Darboe and Kenjale, 2006). The cytoplasm serves as a new niche for *Shigella* survival and replication, until diminished resources prompt invasion of neighbouring cells. The motility of intracellular *Shigella* is supported by IcsA through recruitment of the host's N-WASP and Arp2/Arp3 complex, creating a nucleator site for directional actin polymerization to propel the bacterium through the cytoplasm (Egile, Loisel and Laurent, 1999). Intercellular dissemination occurs via this same mechanism when *Shigella* is endocytosed primarily at tricellular tight junctions, and a new cycle of release-replication-spread takes place.

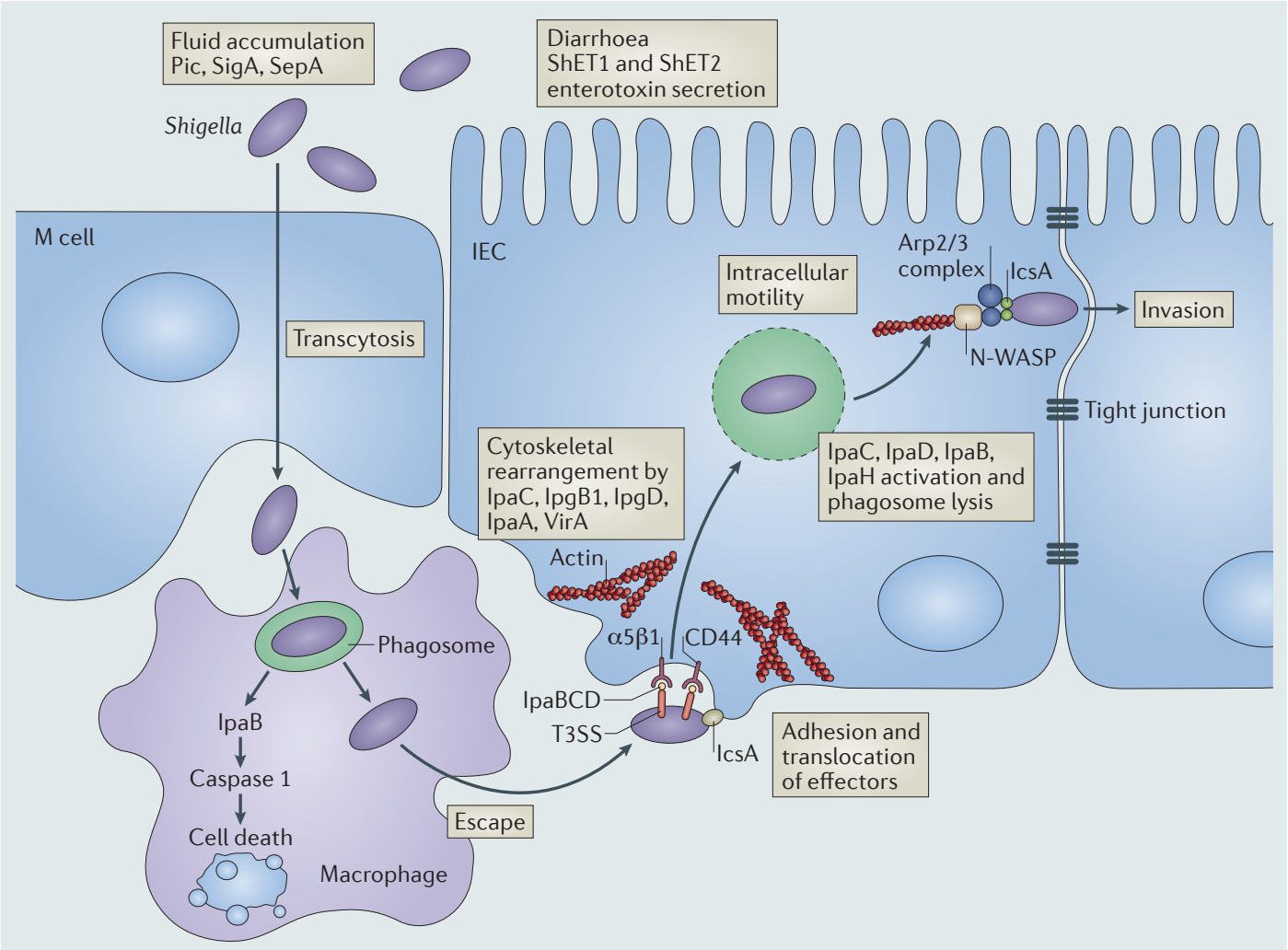


Figure 1.3 The multi-step process of *Shigella*'s molecular pathogenesis (Adopted from (Chung The *et al.*, 2016))

In contrast to the rapid killing of macrophages, *Shigella* must prolong the survival of infected IECs to ensure its own replication. Inside IECs, the secreted IpaB targets Mad2L2, arresting cell maturation (Iwai *et al.*, 2007). Inhibition of cell death and detachment are triggered by activity of IpgD and OspE, respectively (Pendaries *et al.*, 2006; Kim *et al.*, 2009). Cellular secretory functions, including that of pro-inflammatory cytokines, are blocked by disruption of the Golgi and ER-Golgi trafficking by activation of multifunctional IpaB and VirA, respectively (Dong *et al.*, 2012; Mounier *et al.*, 2012). Profuse diarrhoea in *Shigella* infection is attributed to the production of enterotoxins ShET1 and ShET2, which are located in the SHI-1 genomic island and the pINV virulence plasmid, respectively (Nataro *et al.*, 1995; Fasano *et al.*, 1997). In addition, multiple members of the serine protease autotransporter of Enterobacteriaceae (SPATEs) family enhance the bacterium's virulence. The chromosomal genes encoding Pic, SigA and the pINV encoded SepA intensify intestinal fluid accumulation in an animal model, a hallmark of the early stages of the infection (Al-Hasani *et al.*, 2000; Faherty *et al.*, 2012).

1.3.4 The epidemiology of *Shigella*

S. flexneri is currently the major cause of bacillary dysentery in low income settings (in parts of Asia and sub-Saharan Africa, this species accounts for up to 62% of all *Shigella* infections), whereas *S. sonnei* is the most common pathogen in transitional or high-income countries (up to 80% of all *Shigella* infections in Europe and North America are caused by this species) (Gu *et al.*, 2012). *S. boydii* and *S. dysenteriae* cause fewer than 5% each of all cases of shigellosis globally. Notably, *S. dysenteriae* was the main cause of dysentery when it was first identified over a century ago, but today it is infrequently

isolated from patients with dysenteric diarrhoea (Bardhan *et al.*, 2010; Gu *et al.*, 2012). A combination of poor sanitation, malnutrition, unavailability of clean water, and an exceptionally low infectious dose are all potential reasons why *Shigella* is such a successful pathogenic genus and why particular human populations are specifically vulnerable to infection.

Recent epidemiology studies have revealed an intriguing trend in the circulation of different *Shigella* species worldwide. *S. sonnei* is now replacing *S. flexneri* as the most common cause of shigellosis in regions undergoing rapid economic development, where *S. flexneri* had historically dominated as recently as a decade ago (Fig 1.4) (Thompson, Duy and Baker, 2015). This pattern is documented in many countries in Asia, such as Vietnam (Vinh *et al.*, 2009), Thailand (Bangtrakulnonth *et al.*, 2008), and Bangladesh (Ud-Din *et al.*, 2013) as well as in South America and the Middle East (Tajbakhsh *et al.*, 2012). This phenomenon instigates interesting debates underlying its explanatory mechanism and proves to be important to the management and control of shigellosis worldwide. The central hypothesis has revolved around the shared immunity of *S. sonnei* and *Plesiomonas shigelloides,* owing to their almost identical O-antigen structures (Shepherd, Wang and Reeves, 2000). It is conjectured that exposure to *P. shigelloides* serotype 17, found frequently in contaminated surface water, provides immunity to *S. sonnei* via passive 'environmental' immunization (Sack *et al.*, 1994). An improvement in clean water facilities in transitional economies reduces the occurrence of environmental *P. shigelloides*, along with any potential cross-protective immunization. This may explain why *S. sonnei* is able to thrive in these regions. In addition, *S. sonnei* was shown to

invade the ubiquitous amoeba species, *Acanthamoeba castellanii*, *in vitro* and survive within its cytoplasm for more than 3 weeks post invasion (Saeed *et al.*, 2009). This may grant *S. sonnei* protection from environmental stresses and explain its successful circulation.

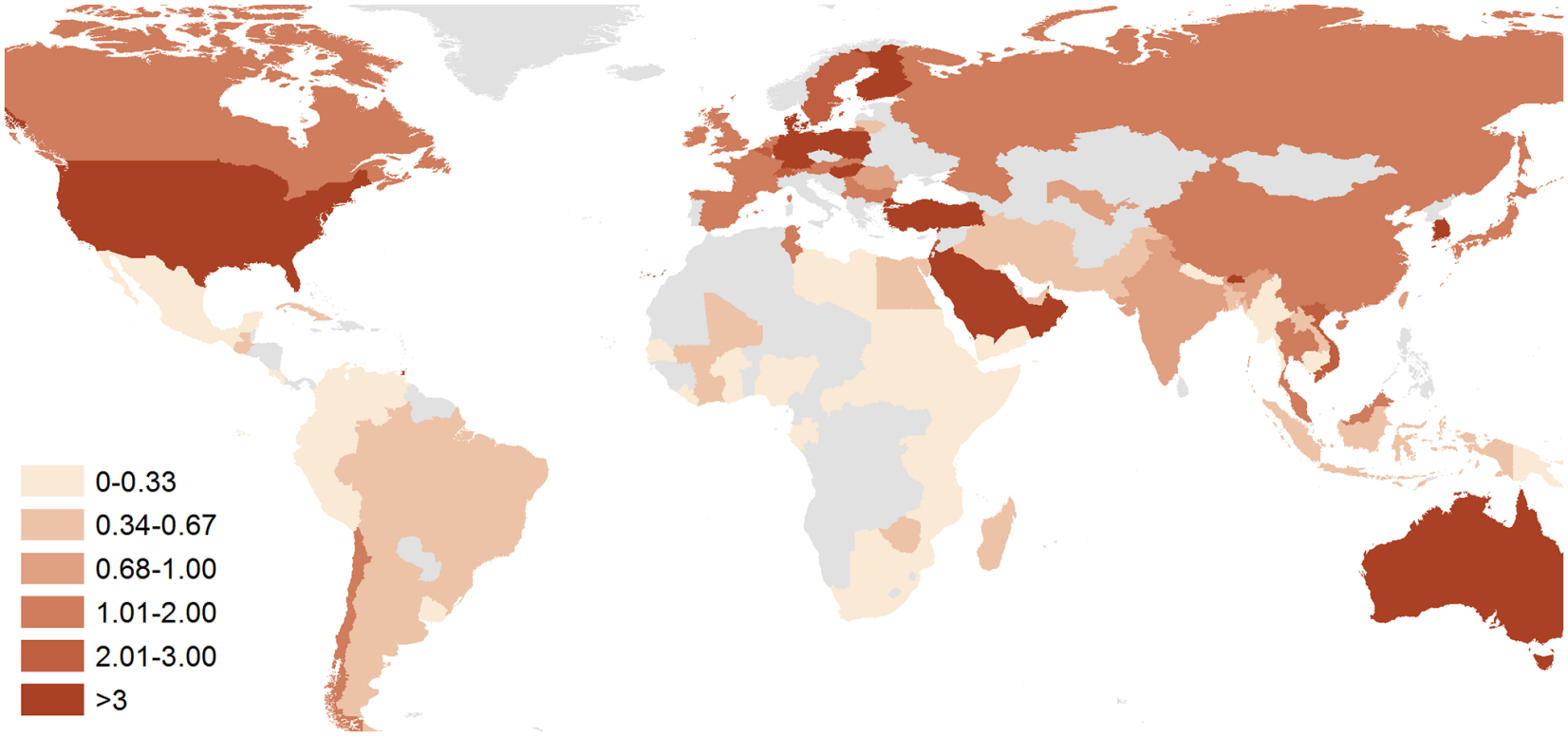


Figure 1.4 Epidemiology of *Shigella* species (Adopted from (Thompson, Duy and Baker, 2015)). The figure shows the ratio of *S. sonnei* to *S. flexneri* from 100 countries (1990-2014), with darker shade attributed to higher ratios (see Legend). Countries coloured grey indicate no data on species identified.

1.4 Shigellosis treatment and management

1.4.1 Treatment of shigellosis and antimicrobial resistance

Shigellosis may result in watery, mucoid and/or bloody diarrhoea due to rupture of the intestinal epithelial cells. Shigellosis management follows the standard guidelines for other diarrhoeal diseases. These include measures to prevent and treat dehydration, such as the use of oral rehydration solution, or supportive care with anti-pyretics and zinc supplementation (WHO, 2005). Although the symptoms are generally self-limiting, with the patients fully recovered within 7-10 days, shigellosis is known to cause debilitating complications in uncommon instances, encompassing metabolic abnormalities, toxic megacolon, intestinal perforation, haemolytic-uraemic syndrome (HUS) and encephalopathy (Bennish, 1991). Notably, shigellosis complicated with encephalopathy leads to more severe symptoms and higher fatality, and it has been recently found to be more associated with *S. sonnei* infection in Bangladeshi children (Afroze *et al.*, 2017). Therefore, antimicrobial treatment is recommended in shigellosis to prevent further complications, as well as to reduce dysenteric discharge and limit post-symptomatic faecal shedding (Vinh *et al.*, 2000; WHO, 2005). However, the choice of antimicrobial has been subjected to debate, and recent data compilations have not been able to conclude the superiority of any certain antimicrobial class (Christopher *et al.*, 2010; Traa *et al.*, 2010; Vinh *et al.*, 2011). Instead, antimicrobial prescribing practice has been constantly updated in response to the circulating pathogen's resistance profile.

Resistance against early generations of antimicrobials has been reported since the 1940s, with the first emergence of sulphonamide resistance in Japan (Niyogi, 2007).

Tetracycline, streptomycin and chloramphenicol were deployed and resistance ensued. During the late 1950s, *Shigella* insensitive to all four aforementioned antimicrobials were prevalent, and this phenotype was later determined to be plasmid-borne (Sköld, 2011). This forced physicians to switch to ampicillin. However, widespread ampicillin resistance in Asia and Africa again diverted shigellosis treatment to another alternative, co-trimoxazole (trimethoprim/sulfamethoxazole), in the 1980s (Niyogi, 2007). Rampant resistance against this drug (co-trimoxazole) was soon observed, with over 60% of *S. dysenteriae* serotype 1 isolates resistant to co-trimoxazole during an Indian dysenteric epidemic in 1984 (Pal, 1984). Therefore, the quinolone nalidixic acid was then deployed to combat the rise of multidrug resistant *Shigella*, and such treatment resulted in satisfactory clinical responses (Bose *et al.*, 1984). Nevertheless, resistance against nalidixic acid rapidly developed and became commonplace in endemic regions by 2000. Besides, the treatment fails to quickly terminate faecal excretion of *Shigella* and is overall more costly in comparison with its successor, the fluoroquinolone (FQ) (Vinh *et al.*, 2000; WHO, 2005). According to the WHO guideline in 2005, FQs such as ciprofloxacin are the first line recommended treatment for shigellosis (WHO, 2005). These quickly become the mainstay treatment for shigellosis as well as acute diarrhoea in endemic countries (Fahad *et al.*, 2010; Kotwani, Chaudhury and Holloway, 2012). Other options include third generation cephalosporins (cefixime) and macrolides (azithromycin). However, growing incidence of resistance against these antimicrobials has been reported. Ciprofloxacin resistant *S. dysenteriae* and *S. flexneri* have been widespread since the early 2000s in India (Taneja and Mewara, 2016), and *S. sonnei* isolated from both endemic and non-endemic settings have shown the predominance of ciprofloxacin

resistance since 2008 (Nandy *et al.*, 2011; Bowen *et al.*, 2015; De Lappe *et al.*, 2015). In Asia and Africa, a recent review estimated that ciprofloxacin resistance in *Shigella* has increased substantially from 0.6% (1998-2000) to 29% (2007-2009) (Gu *et al.*, 2012). As a result, *Shigella* species with resistance to FQ have been listed among the global priority pathogens which urgently demand novel antimicrobial development (Tacconelli *et al.*, 2017). In addition, the circulations of ceftriaxone resistant *S. sonnei* in Asia (Holt *et al.*, 2013) and azithromycin resistant *S. flexneri* among the MSM (men who have sex with men) community (Baker *et al.*, 2015) pose serious challenges in maintaining an effective treatment and management regimen for shigellosis.

1.4.2 Progress in *Shigella* vaccines

The adaptive immune response to *Shigella* infection largely targets its O-antigen (Ferreccio *et al.*, 1991), rendering this structure a sound candidate for vaccine development. However, this is hindered by the wide geographical distribution of a multitude of serotypes, highlighting the need for large-scale surveillance. It was estimated that ~64% of the shigellosis episodes in the GEMS project were caused by only four serotypes: *S. sonnei*, *S. flexneri* 2a, *S. flexneri* 3a, and *S. flexneri* 6 (Livio *et al.*, 2014). Due to the extensive cross protection offered by *S. flexneri* 2a and 3a O-antigens against other serotypes of this species (Noriega *et al.*, 1999), a quadrivalent vaccine composed of these four species' O-antigens could provide ~88% coverage (in the case of 100% efficacious vaccine) against shigellosis (Livio *et al.*, 2014).

Currently, there is no licensed *Shigella* vaccine. Available candidates can be classified into three major approaches: targeting specific O-antigens, common conserved proteins and a combination of both. Live-attenuated variants of *S. sonnei*, *S. flexneri* 2a and *S. dysenteriae* 1, engineered to harbour mutations in essential virulence genes such as *guaBA*, *senAB*, *virG*, or *stxAB*, are entering different phases of clinical study. Furthermore, serotype-specific lipopolysaccharides conjugated with carrier proteins (*Pseudomonas* exoprotein A or tetanus toxoid) are also potential candidates. Purified virulence plasmid proteins IpaB and IpaD were shown to confer protection in animal models, as well as the IcsP and SigA proteins (Martinez-Becerra *et al.*, 2013). Invaplex, a combination of highly conserved *Shigella* IpaBCD and lipopolysaccharide, induces a serotype-specific immune response after intranasal delivery (Riddle *et al.*, 2011).

1.5 Evolutionary history of the genus *Shigella*

The acquisition of the *Shigella* virulence plasmid was the key event in the formation of the different *Shigella* species (Lan *et al.*, 2001), but the origins of this plasmid and the relationship between the species was contentious for a long time. The advent of DNA sequencing and accompanying phylogenetic analysis has permitted a much clearer elucidation of the evolutionary relationships between the different *Shigella* species and their emergence from *E. coli*.

1.5.1 The phylogenetic relationships amongst *Shigella* spp.

Pioneering research in the early genomics era, performed by aligning and comparing the DNA sequences of eight chromosomal housekeeping genes, found three major *Shigella*

clades (C1, C2, C3) and a limited number of outliers, all of which are distinct from, but nested within, lineages of *E. coli* (Pupo, Lan and Reeves, 2000). A further examination of 23 chromosomal genes reached a similar phylogenetic conclusion, albeit with increased resolution, subdividing cluster 1 into three subclusters (SC1, SC2 and SC3) (Yang *et al.*, 2007) (Fig 1.5). Most *Shigella* serotypes from the four species are distributed over the three major clades, demonstrating incongruence between evolutionary history and conventional serology based nomenclature. Cluster C1 contains a combination of serotypes from *S. dysenteriae*, *S. boydii* and *S. flexneri* serotype 6: SC1 includes only *S. dysenteriae* (D3, D4, D6, D9, D11, D12, D13); SC2 contains mostly *S. boydii* (B1, B3, B6, B8, B10, B18) and *S. dysenteriae* serotype 5; SC3 is composed of three other *S. boydii* serotypes (B4, B6, B14) and *S. flexneri* serotype 6 (F6). Major cluster C2 comprises *S. boydii* (B5, B7, B9, B11, B15, B16, B17) and *S. dysenteriae* serotype 2. All *S. flexneri* (F1, F2, F3, F4, F5, FX, FY) except F6 fall into major cluster C3. In this analysis C2 and C3 were found to share a more recent common ancestor than C1, thus emphasizing the close phylogenetic relationship between these two clusters. Analysis of short DNA sequences yielded an estimation of the age of the various clusters, (50,000 to 270,000 years for each of C1 and C2; 35,000 to 170,000 years for C3) (Pupo, Lan and Reeves, 2000). Notable outliers not belonging to any of the three major clusters include *S. sonnei*, *S. dysenteriae* serotypes 1, 8, 10 and *S. boydii* serotype 13 (Fig 1.5). The position of *S. boydii* serotype 13 on the tree topology indicates that it is also distant from the *E. coli*-*Shigella* clade. This genetic distance is consistent with the finding that *S. boydii* 13 and an *Escherichia albertii* group form a discrete lineage, which separated from an *E. coli* ancestor approximately 28 million years ago (Hyma *et al.*, 2005).

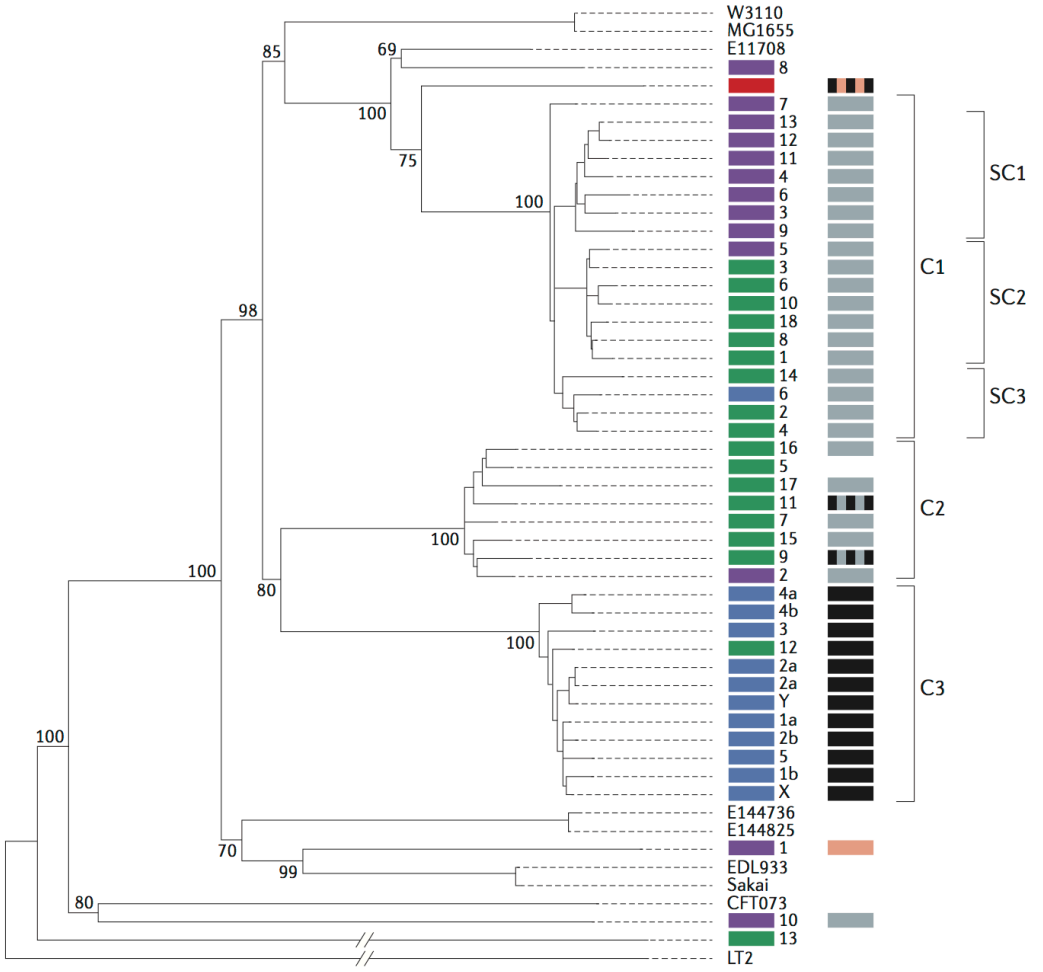


Figure 1.5 The phylogenetic structure of four *Shigella* species (Adopted from (Chung The *et al.*, 2016)). A neighbour-joining phylogenetic tree generated by sequencing 23 chromosomal genes. Strains are labelled by serotype and coloured by species: *S. sonnei* in red, *S. flexneri* in blue, *S. boydii* in green, and *S. dysenteriae* in purple; *E. coli* isolates are uncoloured. Bootstrap values of > 50% are indicated at the major nodes, and the three major *Shigella* genus clusters (C) and subclusters (SC) are indicated. The carriage of the two specific isoforms of virulence plasmids is indicated in the second column: pINV A (grey), pINV B (black), either pINV A or pINV B (hatched black and grey), a unique form of pINV (pink), and either pINV B or a unique isoform (hatched black and pink).

As highlighted above, *S. sonnei* is an outlier from the other *Shigella* species, and the precise phylogenetic relationship between *S. sonnei* and the other *Shigella* species remains ambiguous. It is assumed that *S. sonnei* emerged more recently than the other *Shigella* species and serotypes, with its proposed formation around 10,000 years ago (Shepherd, Wang and Reeves, 2000). However, a more recent study using whole genome sequencing data from globally distributed isolates estimated that all currently existing strains of *S. sonnei* descend from a common ancestor that existed less than 400 years ago, implying that a historical evolutionary bottleneck might have resulted in the extinction of the pre-existing *S. sonnei* strains (Holt *et al.*, 2012).

The early sequence-based genotyping studies described above largely resolved the phylogenetic relationships of different *Shigella* species, yet more recent studies have exploited *Shigella* and *E. coli* whole genome sequences to investigate the evolutionary relationships between these two genera in more detail. Phylogenetic trees for the entire *E. coli-Shigella* group were constructed using an alignment-free feature frequency profile (FFP), which compares genomes based on the frequencies of oligonucleotide sequences with an optimal length for analysis (so called 'features') (Sims and Kim, 2011). These phylogenies, together with those deduced from other studies using core genetic features (present in all genomes and with low variability), have confirmed that the genus *Shigella* is comprised of multiple clusters interspersed within the *E. coli-Shigella* phylogeny, strongly supporting the notion that *Shigella* has emerged from several *E. coli* ancestors on multiple independent occasions (Touchon *et al.*, 2009; Kaas *et al.*, 2012; Zhang and Lin, 2012). The phylogenomic structure of *Shigella* derived from a collection of 336 *E. coli-*

Shigella isolates agrees with the aforementioned grouping based on a limited number of genetic markers as mentioned previously (Sahl *et al.*, 2015). In addition, whole genome resolution phylogenomics also resolves the context for the origins of these major clades: it places C1 and *S. sonnei* in *E. coli* group B1; C2 and C3 in *E. coli* group A; and *S. dysenteriae* 1 in *E. coli* group E (Fig 1.6A) (Sahl *et al.*, 2015). This supports the notion that the phenotypic similarity observed across the *Shigella* species is the result of convergent evolution, in which different *Shigella* founders independently gained genes that facilitate invasive pathogenicity. In the *E. coli* genus, only one pathovar, EIEC, has also acquired invasiveness; EIEC comprises several discrete lineages and exhibits pathogenic and biochemical features that are indistinguishable from those of *Shigella*. Notably, both EIEC and *Shigella* harbour an analogous virulence plasmid, are non-motile and show a negative LDC test (van den Beld and Reubsaet, 2012). These similarities have led to the speculation that EIEC may represent a distinct *Shigella* 'prototype', a precursor for a 'complete' *Shigella* species if selective pressure favours further adaptation of this invasive *E. coli* pathovar (Lan *et al.*, 2004).

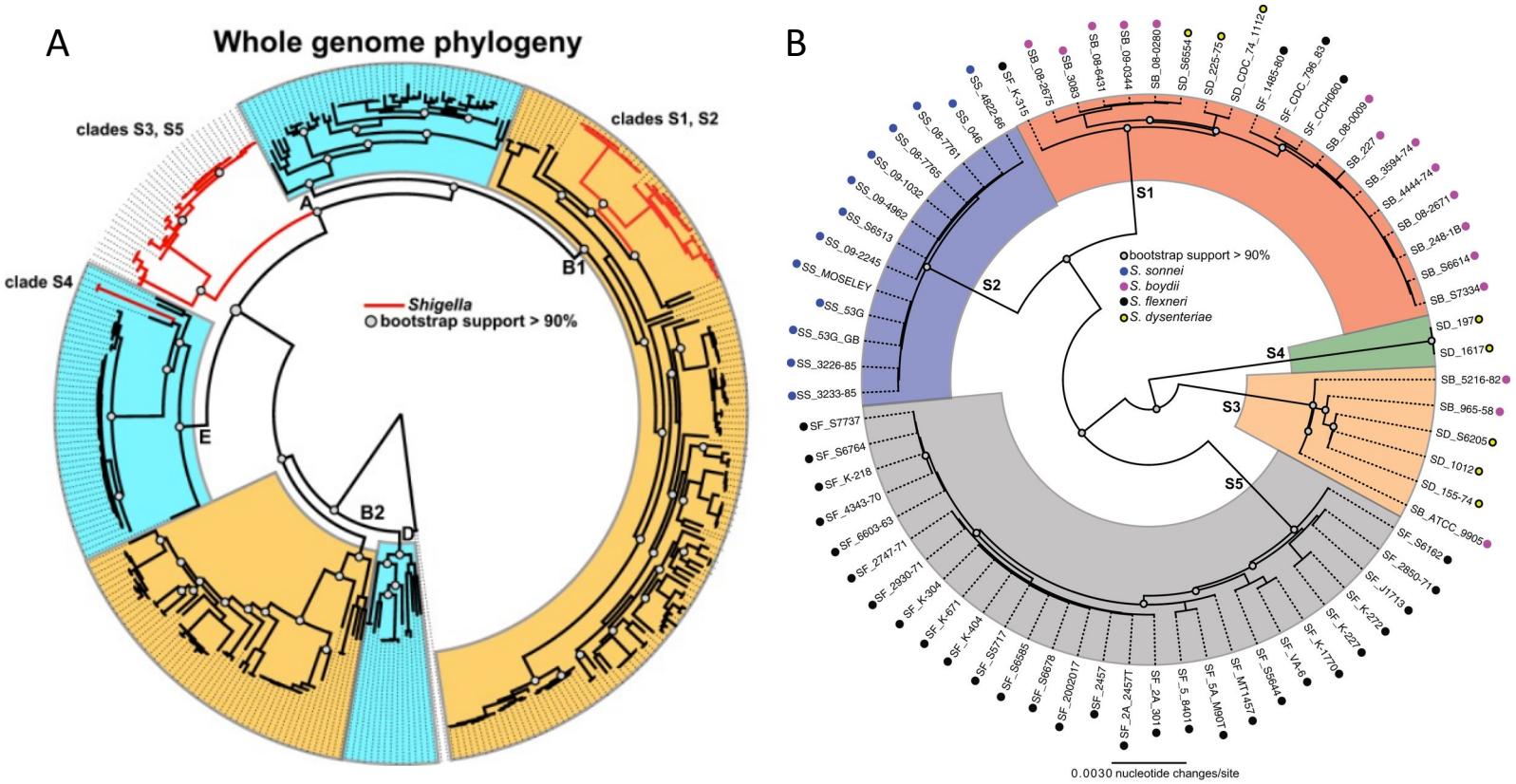


Figure 1.6 Phylogenomics of *Shigella* species (Adopted from (Sahl *et al.*, 2015)). (A) Whole genome phylogeny of *Shigella* in the context of *E. coli* (n=336). The tree was constructed using FastTree2, with 1,000 bootstrap replicates, on the alignment of conserved genomic fragments identified in 40 *E. coli* genomes. (B) Whole genome phylogeny of 69 *Shigella* sequences. The tree was inferred using FastTree2, with 100 bootstrap replicates, on the whole-genome alignment constructed by Mugsy.

1.5.2 Evolutionary history of the *Shigella* virulence plasmid

The *Shigella* virulence plasmid encodes essential virulence factors that facilitate the invasion and spread of *Shigella* into human macrophages and enterocytes, making it a pathogenesis signature (Schroeder and Hilbi, 2008). Various DNA sequencing projects have been performed across several different *Shigella* lineages to elucidate the structure and functions of the virulence plasmid. These projects have uncovered a complex plasmid configuration with a mosaic nature, which is the result of numerous horizontal gene transfer and rearrangement events (Buchrieser *et al.*, 2000; Venkatesan *et al.*, 2001; Jiang *et al.*, 2005) (Fig 1.7). The evaluation of three genes in the *mxi-spa* T3SS region (*mxiA, mxiC* and *ipgD*) revealed two isoforms of the *Shigella* virulence plasmid (pINV A and pINV B) (Lan *et al.*, 2001). Both showed different incompatibility groupings (plasmids of the same incompatibility group cannot be stably inherited in the same cell) (Makino, Sasakawa and Yoshikawa, 1988). When plasmid subtype is mapped onto the *Shigella* phylogeny (Fig 1.5), all C1 isolates harbour pINV A whereas all C3 isolates possess the pINV B isoform. Both forms of the plasmid can be found in C2 strains. The outlier strains harbour either of the two plasmid forms, which is a sign of lateral gene transfer in their history. For example, D10 and most EIEC strains harbour pINV A, whereas *S. sonnei* retains pINV B (Lan and Reeves, 2002; Lan *et al.*, 2004). By contrast, D1 harbours a unique mixed form (pINV A derived *ipgD* and pINV B derived *mxiAC*). This suggests that several ancestral virulence plasmids, from an unknown source, have entered into a diverse background of *E. coli* founder strains. Such introductions include pINV A and pINV B into major *Shigella* clusters C1 and C3, respectively, thus giving rise to these two lineages. Independent acquisitions of either plasmid isoform by *Shigella*

strains not belonging to the main clades as well as lateral gene transfer in C2 complicate inference of the evolutionary history of *Shigella* spp.

Alternatively, plasmid sequences from five virulence genes (*mkaD*, CP0014, *parA*, *parB* and *repA*) located outside the core entry region (defined as the ~30 kbp cluster encoding T3SS and associated effector proteins facilitating the mechanistic invasion into enterocytes) have also been compared. The constructed phylogeny was consistent with the one based on chromosomal genes (Yang *et al.*, 2007), except for a close relationship between C1 and C2 in the plasmid phylogeny, whereas C2 and C3 showed close proximity in the chromosomal phylogeny. The authors argued that the virulence plasmid acquired by C1 and C2 differed from the one obtained by C3. Interestingly, the virulence plasmids from the outliers D1 and *S. sonnei* share considerable homogeneity and can be grouped together outside of the three major clusters. These data suggest that *Shigella* have arisen on multiple independent occasions due to the transmission of multiple virulence plasmid forms into many *E. coli* ancestors. The authors suggested that the subsequent loss of the *tra* locus, which aids the exchange of plasmids between bacteria via conjugation, on the virulence plasmid restricted its transmissibility and enabled parallel evolution of the virulence plasmid and the bacterial chromosomes, thus creating the several discrete *Shigella* lineages we observe today.

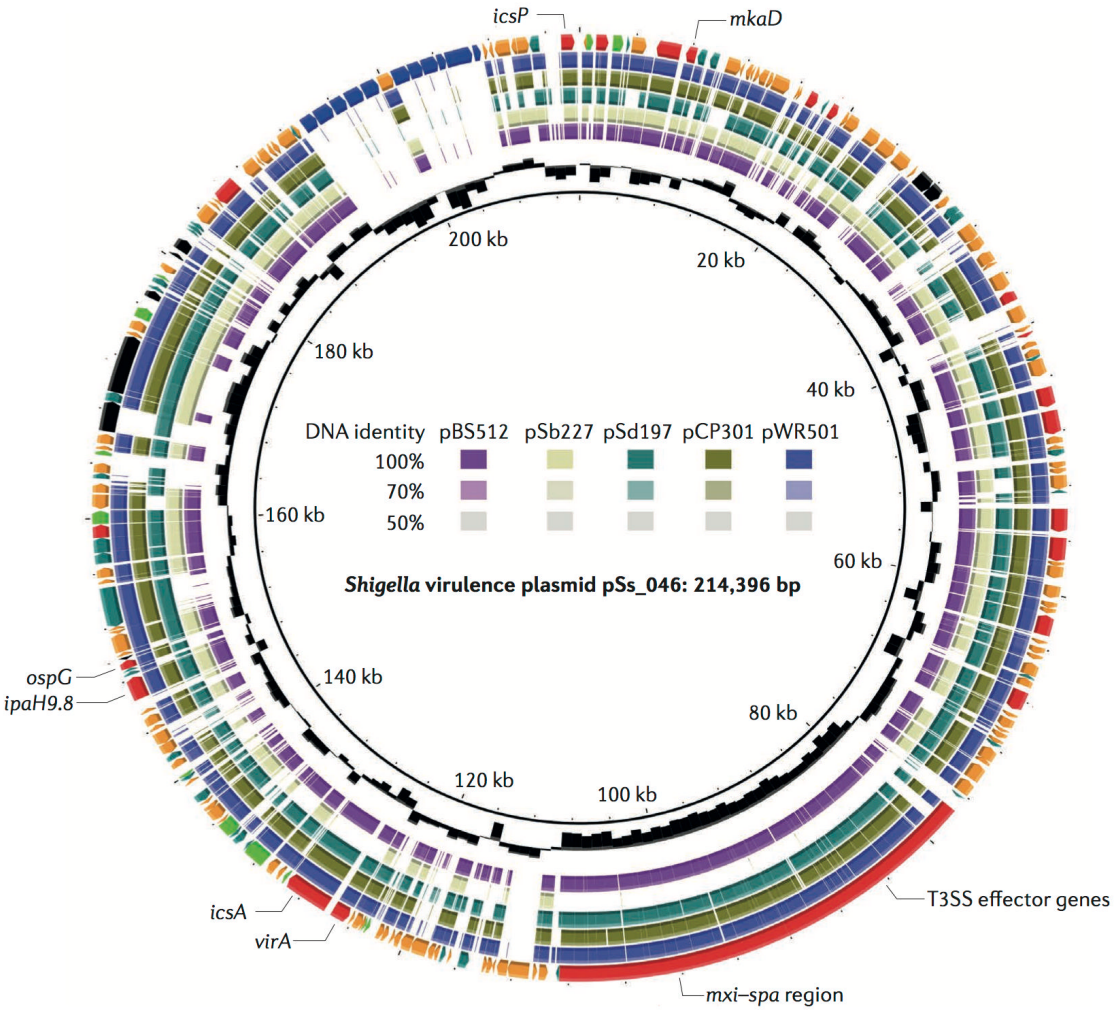


Figure 1.7 A comparative gene map of the *Shigella* virulence plasmid (Adopted from (Chung The *et al.*, 2016)). The innermost ring represents the reference pSs_046, with coordinates. The second ring (black) shows the GC content of pSs_046. The following purple, pale green, teal, khaki, and blue rings show BLASTN comparisons between pSs_046 and the virulence plasmids of *S. boydii* str. BS512, *S. boydii* str. Sb227, *S. dysenteriae* str. Sd197, *S. flexneri* F2a str. 301 (pCP301) and *S. flexneri* F5a (pWR501), respectively. The outer ring represents annotations based on function: known virulence factor genes (red); plasmid replication, transfer and maintenance genes (black); transposon, phage-borne and insertion sequence elements (orange); genes encoding hypothetical proteins (teal); the *S. sonnei*-specific O antigen biosynthesis cluster (blue); and genes encoding proteins with other known functions (green).

1.5.3 Acquisition of genetic elements

The ability to invade host cells and escape the competitive environment of the gastrointestinal tract was pivotal in the emergence of *Shigella* spp. Although acquisition of the virulence plasmid is a 'foothold moment' in the evolution of this pathogen, it is not the only long-term evolutionary change. Numerous other plasmids with different functions have been crucial during *Shigella*'s evolution. For example, an additional small plasmid, pHS-2, harbours a single gene *cld*/*wzz*$_{pHS-2}$, which is instrumental in controlling O antigen chain length and thus leads to resistance to serum killing (Rehel and Szatmari, 1996; Hong and Payne, 1997).

Several clusters of horizontally acquired genetic material, carrying genes that facilitate interactions with the host and contribute towards pathogenesis, have been incorporated into the chromosome of *Shigella*. The largest of these pathogenesis-associated genomic regions are pathogenicity islands (PAIs) and have various functions: encoding an enterotoxin (SHI-1), enabling the sequestration of iron (SHI-2, SHI-3, *sitABCD*), modifying the O-antigen (SHI-O), and conferring resistance to antimicrobials (*Shigella* Resistance Locus - SRL) (Vokes *et al.*, 1999; Al-Hasani *et al.*, 2000; Luck *et al.*, 2001; Ingersoll, Groisman and Zychlinsky, 2002; Schroeder and Hilbi, 2008; Fisher *et al.*, 2009). PAIs have enhanced *Shigella*'s virulence and adaptability, and are commonly associated with bacteriophage integrases, which highlights the major role of bacteriophage in the evolution of *Shigella* spp. One such bacteriophage-associated element is the Shiga toxin (Stx) prophage in *S. dysenteriae* 1; Stx expression can result in severe complications, including haemolytic uremic syndrome (HUS). Recently, an

alternative prophage (φPOC-J13) encoding Shiga toxin stx_{1a} was identified in multiple clinical isolates of *S. flexneri* and *S. dysenteriae* 4 from patients returning from or residing on Hispaniola (Gupta *et al.*, 2007; Gray *et al.*, 2014; M. D. Gray *et al.*, 2015; M. Gray *et al.*, 2015). Unlike the cryptic prophage in *S. dysenteriae* 1, φPOC-J13 appears to be capable of disseminating the stx_{1a} gene into other *Shigella* via transduction (M. Gray *et al.*, 2015). Stx-bearing *S. sonnei* also caused an outbreak in California in 2016, and the *stx* gene was incorporated on a different novel bacteriophage (Kozyreva *et al.*, 2016). Insertion sequence (IS) elements, small transposable DNA sequences capable of 'jumping' within bacterial genomes, are also highly abundant in *Shigella* chromosomes and virulence plasmids. These IS elements have shaped the genomic architecture of *Shigella* spp., causing gene inactivation and genome rearrangement (Buchrieser *et al.*, 2000; Yang *et al.*, 2005; Siguier, Gourbeyre and Chandler, 2014). In relation to genome size, *E. coli* and *Shigella* possess the highest number of IS elements among more than 400 genomes from a range of bacterial species (Wagner and de la Chaux, 2008).

1.5.4 Reductive evolution

Linked to this IS expansion*, Shigella* genomes have also undergone substantial functional gene loss (Ochman and Davalos, 2006). Similar phenomena have been observed in other human-restricted pathogens such as *Yersinia pestis, Mycobacterium leprae* and *Salmonella enterica* subsp. *enterica* serovar Typhi (Eiglmeier *et al.*, 2001; Parkhill *et al.*, 2001; Holt *et al.*, 2008). The modes of gene inactivation are variable in different *Shigella* strains and range from the complete deletion of a locus, to missense point mutations, to IS insertions. However, gene inactivation has occurred preferentially in specific genetic

regions and operons rather than being randomly distributed throughout the genome (Yang *et al.*, 2005; Feng, Chen and Liu, 2011). Independent inactivation of the same or functionally similar genes in different *Shigella* species represents a major pathway of convergent evolution, resulting in similar phenotypic changes, which are associated with adaptation to new niches. For example, different mutations have resulted in the loss of flagellar biosynthesis and specific fimbrial structures (Yang *et al.*, 2005; Bravo *et al.*, 2015). Importantly, loss of flagella results in a reduction in immunogenicity and evasion of the human immune system, as flagellin is a major pathogen-associated molecular pattern (PAMP) (Ramos, Rumbo and Sirard, 2004). In addition, computational reconstruction of metabolic functions (or their losses) based on genomic data groups the different *Shigella* species together and away from other *E. coli* pathovars and commensals, purely by their catabolic functions (Monk *et al.*, 2013). For example, *cadA* and *nadAB* (responsible for the synthesis of cadaverine and the NAD precursor quinolinate, respectively) hinder intercellular spread, phagosomal escape and antigen secretion (Maurelli *et al.*, 1998; Prunier *et al.*, 2007; Prosseda *et al.*, 2012). Likewise, *ompT* and *argT* inhibit intracellular motility and invasive capacity, respectively (Nakata, Tobe and Fukuda, 1993; Zhao *et al.*, 2010). Thus, the loss of these genes in *Shigella* ensures patho-adaptation for an intracellular lifestyle. Alternatively, the loss of gene function can increase the survival rate in a new niche. Disruption of *speG*, which encodes spermidine acetyl-transferase, leads to the accumulation of the polyamine spermidine, which acts as a scavenger of free radicals and thereby provides resistance to oxidative stress in macrophages (Barbagallo *et al.*, 2011). In comparison to *E. coli*, *Shigella* has lost more genes, which is attributed to a reduction in genome-wide purifying selection

and the fixation of inactivating mutations without greatly compromising fitness. It has been suggested that this results from a decrease in effective population size when *Shigella* became a human-restricted pathogen, compared with its *E. coli* ancestor. In addition, the intracellular niche in which *Shigella* began to thrive imposed a more relaxed selective pressure due to abundant resources and a relative lack of competitors (Hershberg, Tang and Petrov, 2007; Balbi, Rocha and Feil, 2009).

1.6 Genomic evolution of *Shigella sonnei*

1.6.1 Global population structure of *Shigella sonnei*

Although *S. sonnei* is the most common *Shigella* species in middle- and high-income countries, the recent emergence of this species in transitional lower-income countries has highlighted the need for more effective surveillance systems and opened new avenues of vaccine research (Wei *et al.*, 2007; Bangtrakulnonth *et al.*, 2008; Vinh *et al.*, 2009; Nygren *et al.*, 2013; Ud-Din *et al.*, 2013). In a key study, the genomes of 132 globally representative *S. sonnei* isolates were sequenced and analysed to investigate the recent evolution of the species (Holt *et al.*, 2012). Three main lineages of *S. sonnei* were identified (I, II, III), which share a most recent common ancestor (MRCA) less than 400 years ago. All lineages likely originated in Europe as the oldest lineages, as well as the majority of all genetic diversity, were detected in European isolates (Fig 1.8). Although all three lineages were distributed in Europe, not all of them had spread globally. Isolates from Asia, Africa and South America were predominantly representatives of the more recently expanded lineage III. Lineage III (particularly clade Global III) emerged in the 1970s and spread internationally in the 1980s and 1990s, establishing more distant

endemic populations in other regions of the world (Fig 1.8). Importantly, there was a correlation between global dissemination and acquisition of resistance to multiple antimicrobials. Resistance was mediated through the gain of class II integrons (In2), which encodes resistance to trimethoprim (*dfrA1*), streptothricin (*sat2*) and aminoglycosides (*aadA1*). The small 8 kbp spA plasmid, conferring resistance to sulfonamide (*sul2*), streptomycin (*strAB*) and tetracycline (*tetA*), was also introduced into multiple *S. sonnei* populations during *S. sonnei*'s evolution. In addition, the pathogen also gained mutations in the DNA gyrase gene, *gyrA*, which encodes the target protein for the broad-spectrum antimicrobials, the FQs. These modifications likely resulted from strong selective pressures induced by antimicrobial exposure.

Recent exploration on *S. sonnei* collected from Latin America and the Caribbean uncovered the existence of lineage V, which is now also detected in South Africa and the UK (Baker *et al.*, 2017). Besides the circulation of the Global III clade in South America (Fig 1.8), this study also reported the expansions of Lineage II (with 2 clades) and another clade belonging to Lineage III within this region. These clades did not show distinctive patterns of temporal or geographic distribution, indicating that different *S. sonnei* populations were constantly transmitted across national borders and intermingled within South America. Independent acquisitions of antimicrobial resistance determinants in these populations, including the introduction of a novel chromosomal SRL island in Lineage III, gave rise to multidrug resistance (MDR) phenotypes (Baker *et al.*, 2017). This indeed may be advantageous in promoting post-symptomatic shedding of *S. sonnei* and sustained short-term transmission in the population (Vinh *et al.*, 2000).

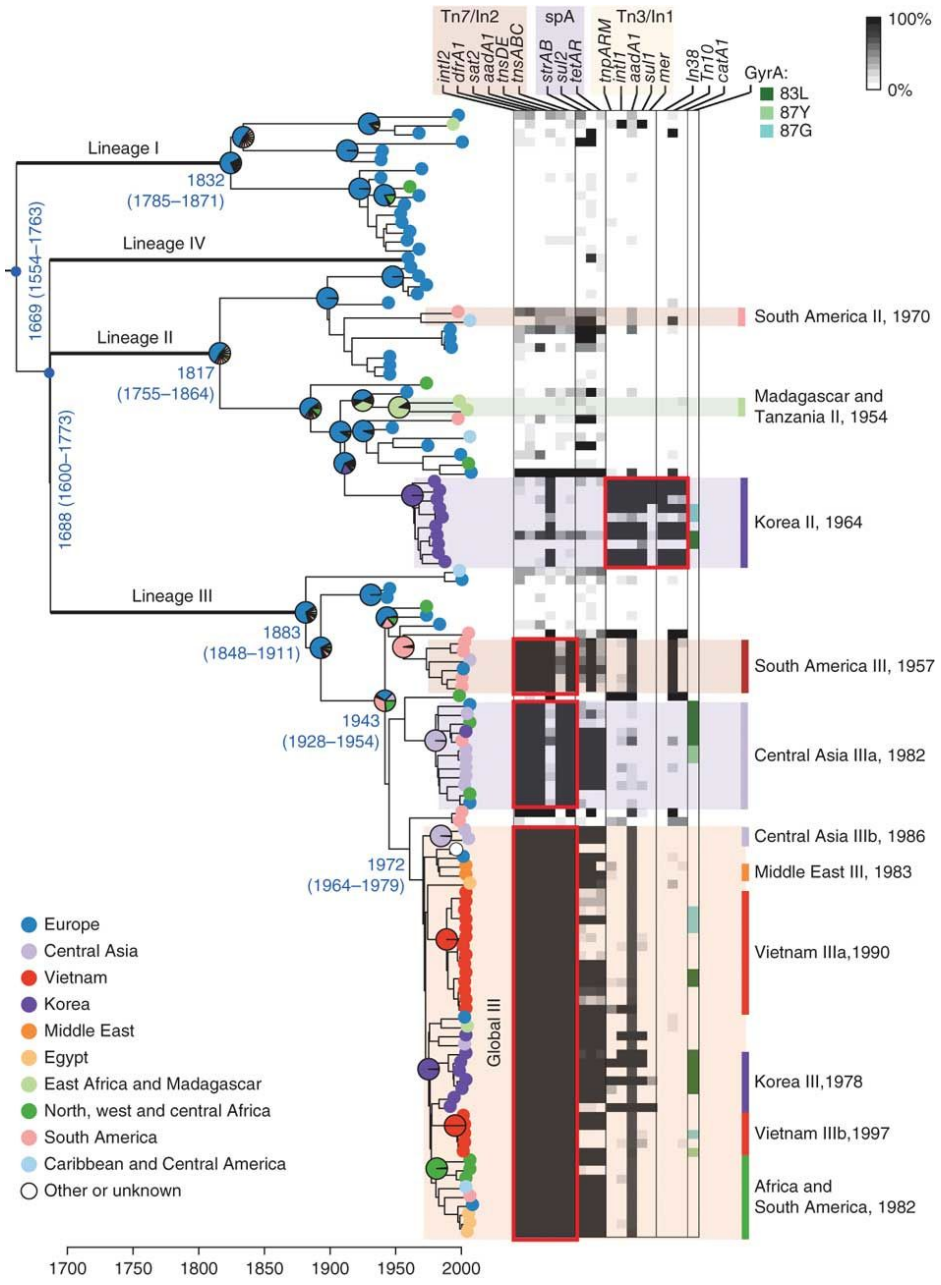


Figure 1.8 Global population structure of *Shigella sonnei* (Adopted from (Holt *et al.*, 2012)). The maximum clade credibility tree is shown on the left, constructed using Bayesian phylogenetics. For major nodes, maximum likelihood estimates for geographic origin are shown as pie charts while divergence dates (median estimates with 95% HPD) are indicated in blue. The heatmap describes the distribution of antimicrobial resistance determinants. Red boxes highlight possible MDR acquisition events.

1.6.2 Evolution of *Shigella sonnei* in Vietnam

A study that investigated more than 250 *S. sonnei* samples in Vietnam expanded upon the observations of the global study (Holt *et al.*, 2013). Genomic and phylogeographical analyses showed that the Global III clade became established in Ho Chi Minh City following the country's reunification in 1975. The founder clone later spread north to other provinces where it established, albeit after multiple introductions, further discrete endemic populations (Fig 1.9). Clonal expansion in these regions contributed to the increase of *S. sonnei*-associated dysentery in Vietnam. In larger human populations, such as in Ho Chi Minh City, a series of bottlenecks in the bacterial population and stepwise accumulation of antimicrobial resistance were co-observed, probably as a consequence of the strong selective pressure exerted by the high use of antimicrobials in the country. Furthermore, plasmid pDPT1, encoding an E5 type colicin (a bactericidal toxin with RNA degradation potency) and an associated immunity protein, became fixed in the Ho Chi Minh City population following the first selective sweep in 1994. This provided a crucial selective advantage over other non-immune *Shigella* and *E. coli* strains. In the 2006 selective sweep, the population acquired plasmid pKHSB1, which harbours an extended spectrum beta-lactamase (ESBL) gene. The acquisition of a plasmid conferring resistance to third generation cephalosporins (Fig 1.9) reoccurred in satellite populations in Khanh Hoa province. Similarly, other signs of convergent evolution included the independent emergences of *gyrA* mutations in Ho Chi Minh City and other provinces, reducing the susceptibility to FQ.

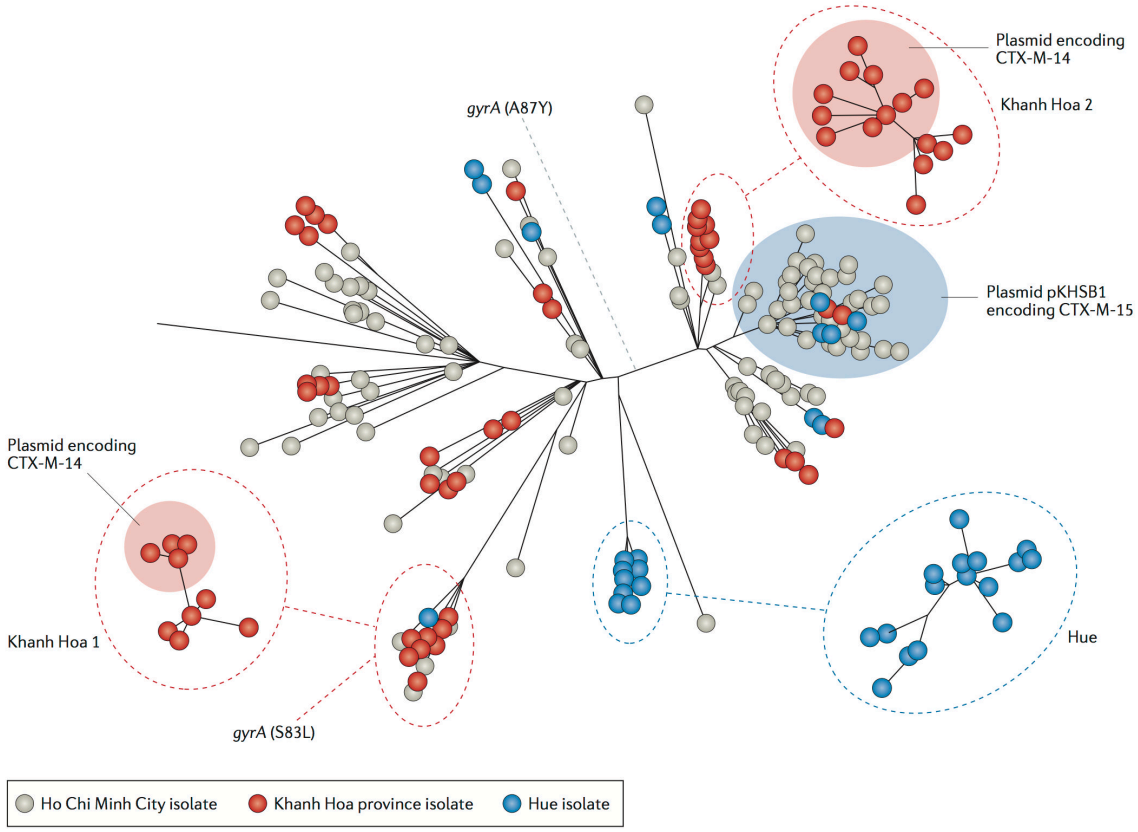


Figure 1.9 Circulation of *Shigella sonnei* in Vietnam (Adopted from (Holt *et al.*, 2013)). Unrooted phylogenetic tree displays the relationship between *S. sonnei* sampled in Ho Chi Minh City (South Vietnam), Hue (Central Vietnam), and Khanh Hoa (South-central coast). Two clonal expansions in Khanh Hoa and one in Hue are the result of the establishment of the founding *S. sonnei* strains after introduction into a specific location. Evidence of homoplasy is demonstrated by the independent emergences of *gyrA* mutations and acquisitions of ESBL plasmids, conferring resistance to quinolones and third generation cephalosporins. Strains harbouring ESBL plasmids are highlighted by coloured shades (blue for incompatibility group I1 (IncI1) plasmid bearing $bla_{\text{CTX-M-15}}$, and red for IncA/C bearing $bla_{\text{CTX-M-14}}$).

1.7 Genomic evolution of *Shigella flexneri*

1.7.1 Genomic determinants of *Shigella flexneri* serotypes

Alongside *S. sonnei*, *S. flexneri* remains the major aetiological agent of bacillary dysentery, particularly in low and middle-income countries. Much of our epidemiological knowledge about *S. flexneri* arises from serotyping data. *S. flexneri* serotypes differ in their O-antigens, and there is experimental evidence for the importance of the O-antigen conformation for invasion and evasion of innate immunity (West *et al.*, 2005). However, serotype conversion (that is, the modification of the serotype in a clonal population) is well documented in *S. flexneri* and mediated by bacteriophages and plasmids, which carry genes contributing to variation of the O-antigen structure. The bacteriophages often integrate as prophage into the chromosomal *thrW* tRNA site (glycosylation operon – *gtr*) or *argW* tRNA site (O-acetylation operon – *oac*), which leads to corresponding modifications in the O-antigen structure (Bastin, Lord and Verma, 1997; Allison and Verma, 2000). Many O-antigen modifying bacteriophages have been identified to date, including SfI, SfII, Sf6, SfIV, SfV and SfX, which convert *S. flexneri* serotype Y into serotypes 1a, 2a, 3b, 4a, 5a and X, respectively (Fig 1.10) (Clark, Beltrame and Manning, 1991; Huan *et al.*, 1997; Mavris, Manning and Morona, 1997; Guan, Bastin and Verma, 1999; Jakhetia, Talukder and Verma, 2013; Sun, R Lan, *et al.*, 2013). Furthermore, the emergence of novel *S. flexneri* serotypes has been widely observed. For example, *S. flexneri* serotype 1c was first identified in Bangladesh in 1980s, and an unrelated clone of this serotype was then also found to be prevalent in rural northern Vietnam and several other Asian countries (Wehler and Carlin, 1988; von Seidlein *et al.*, 2006; Stagg, Cam and Verma, 2008). Moreover, the emergence of *S. flexneri* serotypes 1d, X variant (Xv)

and 4av has been reported in China (Ye *et al.*, 2010; Qiu *et al.*, 2011; Luo *et al.*, 2012). Many of these novel serotypes harbour more than one O-antigen modifying operon, resulting in additional modifications in the already highly modified tetrasaccharide. For example, introduction of *gtrIC* into *S. flexneri* 1a leads to the addition of a glucosyl group on the glucosyl-linked N-acetylglucosamine, effectively converting this serotype into novel serotype 1c (Stagg *et al.*, 2009). Unpredictably, the *gtrIC* cluster shares similarities with proteins from *Citrobacter koseri* rather than with previously characterized orthologs in other *S. flexneri* serotypes. This suggests that *S. flexneri* is able to sample from a large pool of O-antigen modifying genes. Plasmid-mediated serotype conversion has also been reported in *S. flexneri* serotypes Xv, 4av, and Yv. The plasmid-borne *opt* gene was found to be essential for the transfer of phosphoethanolamine (PEtN) to the second rhamnose (RhaII) and the third rhamnose (RhaIII) of the O-antigen in Xv and Yv, respectively (Sun *et al.*, 2013; Zhang *et al.*, 2014).

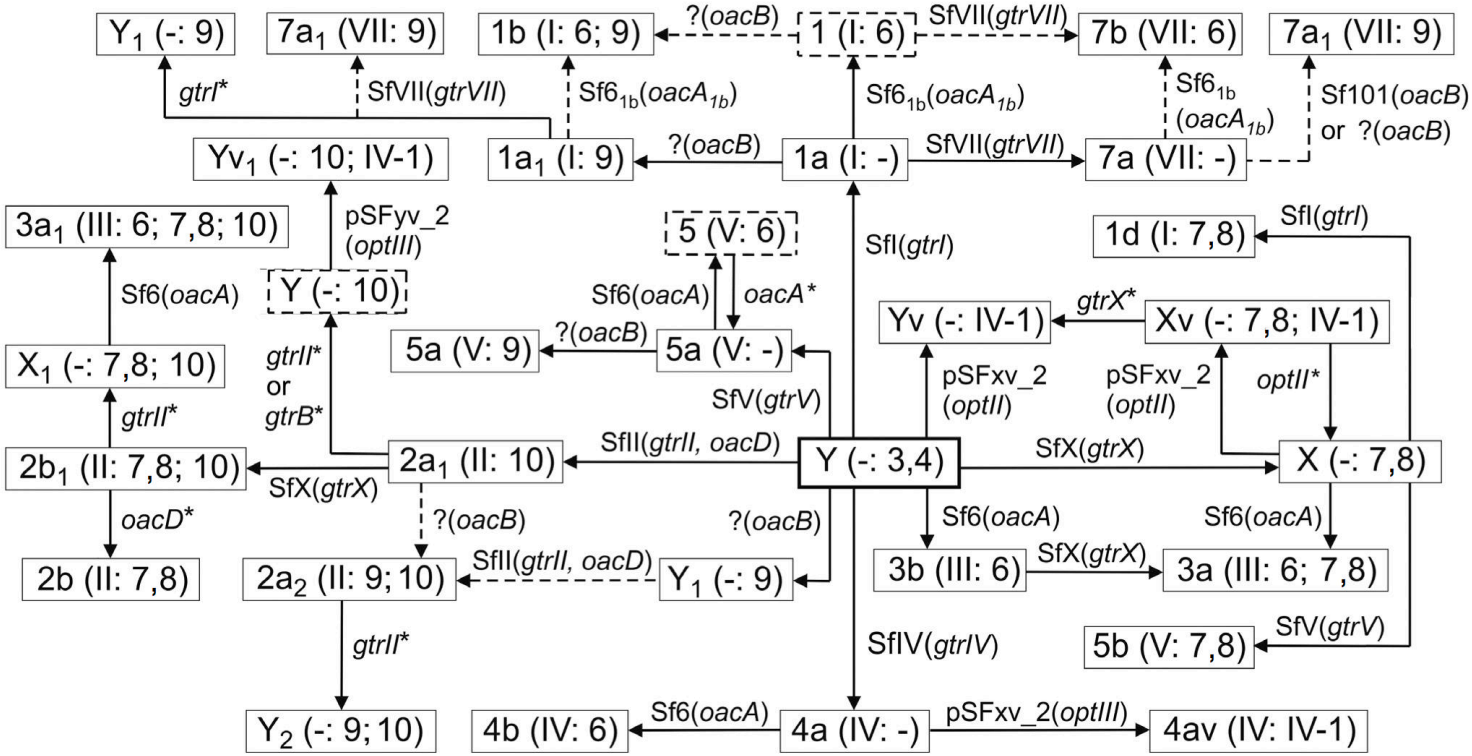


Figure 1.10 Schematic representation of *Shigella flexneri*'s serotype conversion pathway (Adopted from (Knirel *et al.*, 2015)). For each serotype, the antigenic formula is shown within the parentheses. Serotype Y serves as the proposed ancestral serotype, with the O-polysaccharide backbone unmodified. Gene inactivation events are indicated by asterisks. Dashed arrows represent unconfirmed order of serotype-converting events while dashed rectangular boxes indicate hypothetical intermediates not found in nature. Sf61b and SfVII are hypothetical bacteriophages and have not been isolated.

1.7.2 Global population structure of *Shigella flexneri*

Molecular typing of *S. flexneri* has, to date, largely relied on pulsed-field gel electrophoresis (PFGE) and/or multi-locus sequence typing (MLST), using the sequences of the seven housekeeping genes: *adk*, *fumC*, *gyrB, icd*, *mdh*, *purA,* and *recA* (Ribot *et al.*, 2006; Lacher *et al.*, 2007). MLST of more than 100 Asian *S. flexneri* isolates revealed that serotypes 1-5, X and Y belong to a discrete clonal complex (ST245; ST245 complex) whereas serotype 6 forms a distinct clonal complex (ST145; ST243 complex) (Wirth *et al.*, 2006; Choi *et al.*, 2007). MLST does not offer the fine-scaled phylogenetic resolution necessary to investigate the local circulation of specific *S. flexneri* serotypes. However, this approach still holds merit and has been used to examine the spread of the epidemic *S. flexneri* clone ST91 in China (Ye *et al.*, 2010). ST91, typed using another 15 *E. coli*-gene typing scheme, is actually typed ST245 using the *Shigella* MLST approach described above (Lacher *et al.*, 2007). Somewhat atypically for members of the genus *Shigella*, ST91 underwent at least 57 independent serotype switching events during its clonal expansion in China (Ye *et al.*, 2010), illustrating the potential problem in using serotyping as a proxy for genetic relatedness. A major serotype conversion in ST91 led to the rise of a novel variant, serotype Xv, which then rapidly spread and became one of the most prevalent and multidrug resistant serotypes in China since 2000 (Ye *et al.*, 2010). Whole genome sequencing provides much higher phylogenetic resolution, and its application showed that ST91 serotype Xv had acquired plasmids carrying *opt*, leading to O-antigen modification, on three independent occasions (Zhang *et al.*, 2014). Before the *opt* harbouring plasmid was introduced, ST91 had already carried antimicrobial resistance genes, including the MDR SRL (a genomic island harbouring resistance genes

against tetracycline (*tetACDR*), streptomycin (*aadA2*), ampicillin (*oxa-1*) and chloramphenicol (*cat*)), Tn7 (an MDR transposon carrying resistance genes against trimethoprim (*dfrA1*), streptothricin (*sat1*) and streptomycin (*aadA1*)), and two mutations in *gyrA* facilitating resistance against nalidixic acid. ST91's rapid expansion in different geographical locations can be explained by O-antigen switching and evasion of pre-existing immunity in populations.

In addition to the dramatic species shift observed in developing countries, *S. flexneri* epidemiology in certain populations in developed countries has also changed. The isolation rate of *S. flexneri* 3a has increased steadily in MSM communities in Canada, England and Wales (Borg *et al.*, 2012; Ratnayake, Allard and Pilon, 2012). This increase is not attributable to an introduction(s) from the low-income countries, suggesting that the ecology of this particular variant may now be better adapted to transmission within MSM populations (Borg *et al.*, 2012). A recent genomic analysis of a global collection of this serotype revealed the emergence of a *S. flexneri* 3a lineage attributed to infections in MSM population in higher-income countries (Baker *et al.*, 2015). This lineage has spread globally since its emergence in 1998, and as common for contemporary populations of *Shigella*, has acquired resistance against multiple antimicrobials, most notably against azithromycin, a frequent antimicrobial treatment for sexually transmitted diseases (STIs) including gonorrhoea, syphilis and chlamydia. This change in antimicrobial susceptibility, seen as the response to selective pressure exerted by azithromycin treatment for other comorbid infections, may have contributed to the dominance of this organism in MSM (Baker *et al.*, 2015).

Studying *S. flexneri*'s evolution and epidemiology proved complicated due to serotype diversity until a recent study of 351 whole genome sequences from different serotypes of this species (Connor *et al.*, 2015). This study concluded that *S. flexneri*, with the exclusion of the diverging serotype 6, consists of seven phylogenetic groups (Fig 1.11). Notably, these phylogenetic groups are inconsistent with serotype-based groupings and have arisen on multiple occasions between the 1300s and the 1800s (Connor *et al.*, 2015). The presence of numerous serotypes in all phylogenetic groups suggests that serotype switching is commonplace, consistent with previous research (Ye *et al.*, 2010). This study also revealed substantial variability in the composition of *S. flexneri* virulence factors (e.g., the SHI-1 genomic island, iron uptake systems such as enterobactin and the *fec* locus) and antimicrobial resistance genes (e.g., the SRL island). SHI-1, SRL and enterobactin exclusively co-occur in phylogenetic group 3, which is comprised predominantly of serotype 2a, and this may account for the enhanced virulence and international dominance of this serotype (Connor *et al.*, 2015). The accumulation of antimicrobial resistance genes in *S. flexneri* over the past three decades is considered to be essential for maintaining successful lineages. However, unlike for *S. sonnei*, this has neither led to displacement of pre-existing antimicrobial susceptible lineages nor dramatic international transmission, with the exception of the global spread of the MSM-associated *S. flexneri* 3a (Holt *et al.*, 2012; Baker *et al.*, 2015). This supports the concept of longer-term colonization, where diverse populations of both antimicrobial resistant and antimicrobial susceptible lineages co-circulate in endemic locations. These data also imply that *S. flexneri* may be persisting in the environment, where selection for antimicrobial resistance may be less favourable.

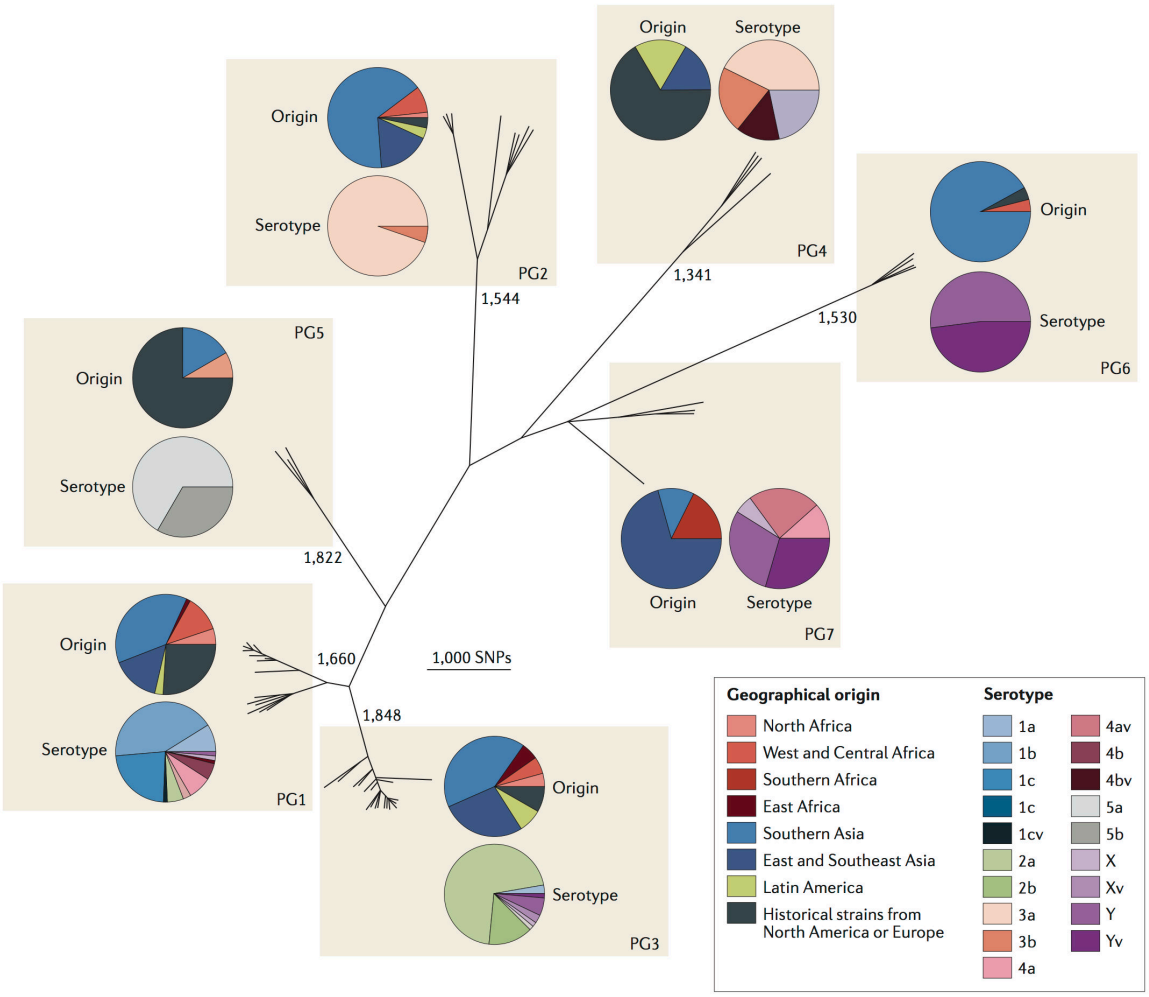


Figure 1.11 The global phylogenetic structure of *Shigella flexneri* (Adopted from (Chung The *et al.*, 2016)). The figure displays a maximum likelihood phylogeny of *S. flexneri* serotypes from a global collection of 351 representative isolates, including serotypes 1-5, X, Xv, Yv, and Y. Phylogenetic groups (PGs) were identified using Bayesian analysis of population structure clustering, and the geographical and serotypic distribution from each PG are summarized as respective pie charts.

1.8 Genomic evolution of *Shigella dysenteriae* and *Shigella boydii*

As *S. dysenteriae* and *S. boydii* account for fewer than 10% of the cases of shigellosis, research into these organisms is less of a priority for global health research (Gu *et al.*, 2012). Furthermore, research is complicated by their sheer diversity in serotypes (15 for *S. dysenteriae* and 20 for *S. boydii*) and a lack of large, well-characterized, geographically diverse collections of isolates.

One of the best studied *S. dysenteriae* serotypes is serotype 1 (*Sd*1), which induces a more severe disease phenotype than other *Shigella*. The hypervirulence of *Sd*1 can be explained by the release of Stx, which inhibits protein synthesis (Khan, Griffiths and Bennish, 2013); also potentially contributing to this hypervirulence is the presence of the *shu* cluster, which is upregulated in response to the host's body temperature and uses haem as an iron source, leading to better adaptation in the human host (Wyckoff *et al.*, 1998; Kouse *et al.*, 2013). Notably, enterohaemorrhagic *E. coli* (EHEC) O157:H7 also carries the *stx* prophage and the *shu* cluster and can also cause severe complications, demonstrating that *Sd*1 and EHEC have inherited and maintained these virulence factors from a common ancestor (Wyckoff *et al.*, 1998). The last reported dysentery outbreak caused by *Sd*1 occurred in Sierra Leone in 1999 (Guerin *et al.*, 2003); since then, the prevalence of disease due to this serotype has become negligible (Bardhan *et al.*, 2010). A detailed study of more than 300 temporally and spatially diverse *Sd*1 whole genome sequences proposed the existence of four *Sd*1 lineages, with their divergence dated back to the 18th century (Njamkepo *et al.*, 2016). Intercontinental transmissions heightened quickly during the 20th century, and recent waves of introductions from South Asia to

Africa have resulted in multiple epidemics. The rapid dissemination of *Sd*1 was facilitated by poor sanitation and intense human migration during the 19th century, and period surrounding the two world wars (Rohmer *et al.*, 2014; Njamkepo *et al.*, 2016). Furthermore, these *Sd*1 lineages independently acquired several antimicrobial resistance genes, particularly the accumulation of the SRL-PAI genomic island (carrying resistance to ampicillin, streptomycin, chloramphenicol, and tetracycline) and Tn7 transposon (encoding trimethoprim resistance) in lineage IV. Resistance to FQ was also evident in this *Sd*1 lineage since 1995, making its MDR status a significant public health threat. This, together with its unexplained significant decrease in isolation rates globally, warrants further research for proactive epidemic dysentery management.

S. boydii was first isolated on the Indian subcontinent and appears to be restricted to this region, as it is rarely isolated elsewhere (Khan *et al.*, 1985). However, a new serotype, *S. boydii* serotype 20, was discovered in travellers to Central America, which demonstrates that the epidemiology of *S. boydii* is more complicated than previously described or predicted (Kalluri *et al.*, 2004; Woodward *et al.*, 2005). In developing countries, serotype 2 is the most prevalent and clinically relevant serotype, with an isolation rate around 50% of all *S. boydii* isolates (Smith *et al.*, 2009; El-Gendy *et al.*, 2012; Livio *et al.*, 2014). Other *S. boydii* serotypes are rare, but several O-antigen clusters from *S. boydii* have been transferred to different members of the genus *Escherichia*; for example, *S. boydii* O-antigens 10 and 15 can be found in Stx producing *E. coli* (STEC) and *Escherichia fergusonii*, respectively (Iguchi *et al.*, 2011; Azmuda *et al.*, 2012).

1.9 Aims and objectives of the thesis

Despite *Shigella* being an important diarrhoeal pathogen worldwide, its epidemiology and evolution remain incompletely characterized in endemic regions, especially in Asia. This DPhil thesis sets out to use whole genome sequencing and state-of-the-art genomic and phylogenetic analyses to understand *Shigella* evolution in multiple settings and aspects. These include the following primary objectives:

(1) To investigate the unusual surge of FQ resistant *S. sonnei* in Bhutan during 2011-2013.

(2) To unravel the nature of FQ resistant *S. sonnei* at the global scale, and to characterise the epidemiology and evolution of this major clone in detail.

(3) To examine the evolution and transmission of both *S. sonnei* and *S. flexneri* in Southeast Asia during the past three decades, utilizing a vast collection of historical sequences.

A secondary objective of this thesis was to use 16S rRNA sequencing to understand the dynamics of the human gut microbiome during infectious diarrhoea, with aetiologies ranging from viruses to bacteria such as *Shigella*. This study demonstrates the versatility of applying sequencing to studying diarrhoea in endemic regions and adds new insights to the current limited literature on the infectious diarrhoeal gut microbiome.

2 Materials and Methods

2.1 Methods applicable throughout chapters 3 to 5

2.1.1 Whole genome sequencing (WGS) of *Shigella* isolates

Unless specified otherwise, the procedure for DNA extraction and sequencing is as described below. Total genomic DNA was extracted from all confirmed *Shigella* isolates using the Wizard Genomic DNA Extraction Kit (Promega, Wisconsin, USA), following the manufacturer's instructions. DNA extraction was performed by staff scientists at OUCRU or collaborating institutions. Extracted DNA was stored at -80°C until shipment to the Wellcome Trust Sanger Institute (WTSI) in Hinxton, United Kingdom, for sequencing. For each sample, 2μg of genomic DNA was used for WGS on an automated Illumina Hiseq2000 platform to produce paired-end reads of length either in 100 or 125 bp. Sequencing reads were then archived and processed by WTSI's Pathogen Informatics Quality Control (QC) pipeline, which maintains a stringent quality assessment of sequencing data.

2.1.2 Mapping sequencing reads to references and detection of single nucleotide polymorphisms (SNPs)

For each read set, sequences were mapped to a suitable chromosomal reference, Ss046 (accession number NC_007384) for identified *S. sonnei* and Sf2a strain 301 (accession number NC_004337) for identified *S. flexneri* isolates respectively. Unless indicated otherwise, the mapper of choice is SMALT v0.7.4 , which employs hash-indexing algorithm to search for exact matches of reads to the indexed reference (Ponsting and Ning, 2010). References were indexed with the choices of k-mer sizes of 13 (-k 13) and

sampling at every sixth position in the genome (-s 6) for read length equal to 100 bp; and k-mer sizes of 20 and sampling at every 13th position of the genome for read length equal to 125 bp. Mapping quality was individually assessed, using the average coverage depth and mapping percentage to reference as key evaluators. Candidate SNPs were called against the reference sequence and filtered using SAMtools (mpileup option) and bcftools, respectively (Li *et al.*, 2009). Low quality SNPs were removed according to the following criteria: consensus quality <50, mapping quality <30, ratio of SNPs to reads at a position <75%, read depth <4, number of reads per strand <2. This helped generate a high-quality SNP set for each sample. A pseudo-sequence, with length identical to the mapped reference, was constructed for each sample to incorporate high-quality SNPs, genomic regions of low mapping coverage, and invariant sites. However, these pseudo-sequences exclude genomic regions which are absent from the references, extra-chromosomal or specific to each isolate. This makes pseudo-sequences suitable for phylogenetic inference. All these procedures were carried out in the WTSI's Pathogen Genomic Mapping and SNP Calling pipeline, which allows rapid and efficient high-throughput SNP detection.

2.1.3 Phylogenetic reconstruction

Pseudo-sequences of samples of interest were concatenated to generate an alignment. Genetic elements contributed by horizontal gene transfer (HGT), which include recombination regions and mobile genetic elements such as prophage and genomic islands, do not represent vertical evolution. Failure to exclude these elements proves to lead to inaccurate inference, concerning both the tree's topology and branch lengths.

Gubbins, which uses spatial scanning statistic to detect genomic regions with significantly high SNP density, was used on the alignment to remove potential recombination regions (Croucher *et al.*, 2014). Other genomic elements annotated as prophages, insertion sequences (IS), and genomic islands in the references (Ss046 and Sf2a strain 301) were also removed from the alignment using the script 'remove_blocks_from_aln' (https://github.com/sanger-pathogens/remove_blocks_from_aln). Invariant sites were then filtered from the processed alignment to generate a SNP-only alignment.

Unless specified otherwise, phylogenies were inferred using maximum-likelihood method on the prepared SNP-only alignment. The method is implemented in RAxML v8.1.3 with the General Time Reversible (GTR) model, and sufficient bootstrap replicates were automatically determined by the extend majority rule (MRE) bootstrap convergence (Stamatakis, 2014). The GTR is the most general model of DNA substitution, which allows unequal frequencies for four nucleotide bases and six unequal nucleotide substitution rates (Tavaré, 1986). For each alignment, appropriate most closely related outgroups were also included if possible to generate a high confident rooted phylogeny.

2.1.4 Temporal analysis and Bayesian phylogenetic inference

When the time of isolation was recorded for *Shigella* isolates, it is of interest to investigate the time to most recent common ancestor (tMRCA) of the population as well as its whole evolutionary history. Maximum likelihood phylogenies of the selected subset (including a fixed maximum number of isolates per calendar year to avoid biased

sampling) were constructed as mentioned in section 2.1.3. Path-O-Gen (v1.4) or TempEst was used to estimate the linear relationship between root-to-tip divergence of the input maximum likelihood phylogeny and the isolates' sampling date (in month/year) (Rambaut *et al.*, 2016). Only phylogenies with sufficient temporal structure, as assessed visually or by R^2 statistics, were further used for Bayesian phylogenetic inferences. This ensures the examined population possesses sufficient genetic change over the sampling period, which satisfies the assumption of the existing molecular clock hypothesis. The most appropriate substitution model for the input alignment was determined using either jModeltest or the fast ModelFinder implemented in iqtree (Posada, 2008; Nguyen *et al.*, 2015; Kalyaanamoorthy *et al.*, 2017).

Bayesian Markov Chain Monte Carlo (MCMC) deployed in BEAST (Bayesian Evolutionary Analysis of Sampling Trees; v1.8.0 and v1.8.3) was then used to infer the evolutionary dynamics of the input date-stamped SNP-only alignment (Drummond and Rambaut, 2007). To identify the best-fit models for the dataset, multiple BEAST runs were conducted on combinations of the most suitable substitution model (usually TVM + Γ_4), a clock model, and a demographic model. The clock model is comprised either of a strict, uncorrelated relaxed lognormal, or random local (in some cases) model (Drummond *et al.*, 2006; Drummond and Suchard, 2010). The demographic model includes either a constant, or piecewise Bayesian skyline model (Drummond *et al.*, 2005). Each of these analyses was performed in triplicates using a continuous 50 to 200 million generation MCMC chain with samples taken every 5,000 to 20,000 generations. This generates an output of 10,000 trees for each BEAST run. Parameter convergence was

assessed visually in Tracer v1.5, indicated by effective sample size (ESS) values > 200. For robust model selection, both path sampling and stepping-stone sampling approaches were applied to each BEAST analysis to estimate the marginal likelihood (Baele *et al.*, 2012, 2013), and the best model was selected based on comparison of Bayes factor. For the best model, triplicate runs were combined by LogCombiner v1.8.3, with removal of 20% burn-in. Calculated parameters from the chosen model were reported and discussed in analyses.

2.1.5 Construction of de novo assemblies

Unless indicated otherwise, sequence assembly for each isolate was performed according to the procedure described by the WTSI's Pathogen Genomic Bacteria Assembly pipeline, which facilitates robust and improved high-throughput prokaryote assembly (Page *et al.*, 2016). First, de novo sequence assembly was constructed using Velvet v1.2.03, with the optimal size of k-mers determined by VelvetOptimizer by varying its size between 66% and 90% of read length (Zerbino and Birney, 2008). Velvet allows effective manipulation of de Bruijn graphs, which are graphical representation of sequenced data based on k-mers, to restrict mis-assembly errors and resolve repeats. To improve assembly quality, SSPACE was run in 16 iterations to scaffold the contiguous sequences (contigs) (Boetzer *et al.*, 2011), and GapFiller was performed in 120 iterations to fill in sequence gaps (Boetzer and Pirovano, 2012). Contigs below 300 bp were excluded, and read pairs were mapped back to the improved assembly using SMALT to assess the quality of assembly. Assembly performance was assessed for each sample, using the number of contigs and N50 length (contig length such that half of the

nucleotides in the entire assembly lie in contigs equal or larger than this value). Each assembly was then annotated using PROKKA, rapidly producing rich and reliable annotations for bacterial genomes (Seemann, 2014). Due to the omnipresence of insertion sequences (IS) and other repetitive elements in *Shigella* genomes (Yang *et al.*, 2005), assembly algorithm fails to resolve repeats, leading to the composition of hundreds of contigs in one *Shigella*'s de novo assembly.

2.1.6 Determination of the accessory genome

To identify contigs associated with the accessory genome, contigs were first ordered with the reference (*S. sonnei* Ss046 or *S. flexneri* Sf2a301) and its large virulence plasmid (pSS046 or pCB301) using ABACAS (Assefa *et al.*, 2009). The remaining contigs were visualized and examined in Artemis. Presence of specific accessory genetic determinants was confirmed by BLAST and visualized using Artemis Comparison Tool (ACT) (Carver *et al.*, 2008).

ARIBA, a platform to detect genetic elements of choice from short read sequences, was deployed to detect the presence of antimicrobial resistance (AMR) genes, plasmid incompatibility types, and colicin types based on the reliably curated ResFinder, PlasmidFinder, and an in-house colicin database, respectively (Zankari *et al.*, 2012; Carattoli *et al.*, 2014; Hunt *et al.*, 2017). ARIBA outputs from each isolate were then compiled and curated. Genes that contained deletion, frameshift or nonsense mutations were deemed non-functional and counted as being absent from the isolate.

2.1.7 Stochastic mapping of geographical traits

For a *Shigella* population of interest, I aimed to reconstruct the ancestral geographical state of each isolate, using the reference-based maximum likelihood phylogeny as an input. The method was adopted from a recently described approach to investigate the geographical spread of *Burkholderia pseudomallei* in Southeast Asia (Chewapreecha *et al.*, 2017). I treated geographical source (at different levels ranging from cities to subcontinents) of the organisms as discrete characters. To control for biased sampling, I sub-sampled the input phylogeny to include an equal number of isolates (N) from each character, generating 500 subsampled trees. Geographies with fewer than N isolates were excluded in this analysis. I utilized stochastic mapping to quantify transition events between geographical characters as well as the total time spent within each character state, separately for each subsampled tree (Bollback, 2006). Stochastic mapping, implemented as the function *make.simmap* in the R package phytools v0.6.0, was performed under an asymmetric model of character change (ARD) with the rate matrix sampled from the posterior probability distribution using MCMC (Q=mcmc) for 100 simulations (Revell, 2012).

2.2 Methodology specific to Chapter 3: The introduction and establishment of fluoroquinolone resistant *Shigella sonnei* into Bhutan

2.2.1 Sample collection

The Research Ethics Board of Health in Bhutan and the Institutional Review Board of the Walter Reed Army Institute of Research provided ethical approval for the analysis carried out in this study. Stool samples were collected during a diarrhoea surveillance study

conducted at the Jigme Dorji Wangchuk National Referral Hospital (JDWNRH) in Thimphu, Bhutan, between March 2011 and October 2013, as previously described (Ruekit *et al.*, 2014). Briefly, children aged between three months and five years presenting to JDWNRH with acute diarrhoea (three or more loose stools in the previous 24 hours, with symptoms lasting ≤72 hours at time of recruitment) and those with no diarrhoeal symptoms in the previous two weeks were eligible for recruitment as cases and controls, respectively (informed consent permitting).

2.2.2 Bacteriology and antimicrobial susceptibility testing

Stool samples were collected from cases and controls and cultured on MacConkey and Hektoen agar and incubated overnight at 37ºC. Non-lactose fermenting colonies were confirmed to be *Shigella* by standard biochemical testing; serotype was identified by O-antigen agglutination (Denka Seiken, Japan). All *Shigella* isolates were subjected to antimicrobial susceptibility testing using the disk diffusion method as recommended by CLSI guidelines (CLSI, 2012). The following antimicrobials were selected for susceptibility testing: azithromycin (AZM 15 μg), nalidixic acid (NA 30 μg), ciprofloxacin (CIP 5 μg), ampicillin (AMP 10 μg), trimethoprim-sulfamethoxazole (SXT 1.25/23.75 μg), ceftriaxone (CRO 30 μg), streptomycin (STR 10 μg), and tetracycline (TET 30 μg). All laboratory work pertaining to this study was performed by researchers at the Armed Forces Research Institute of Medical Sciences (AFRIMS), Bangkok, Thailand.

2.2.3 Whole genome sequencing

During the surveillance study, a total of 112 *Shigella* were isolated (74 *S. sonnei*, 30 *S. flexneri*, 7 *S. boydii* and 1 *S. dysenteriae*); the 74 *S. sonnei* were selected for WGS. The majority (64/74) of the selected *S. sonnei* were isolated from symptomatic diarrhoeal stool samples with the remaining strains isolated from asymptomatic infections.

Genomic DNA was extracted from all *S. sonnei* and sent to the WTSI for sequencing. For all samples, reads were mapped to the reference sequence of *S. sonnei* Ss046 (accession no. NC_007384) using SMALT (version 0.7.4), and high quality SNPs were detected as described in section 2.1.2. Three samples failed due to insufficient read depth, resulting in a final sample set including 71 *S. sonnei* whole genome sequences from Bhutan.

2.2.4 Phylogenetic reconstruction of Bhutanese *S. sonnei*

In order to place the Bhutanese isolates in the context of global diversity, I aligned the Bhutanese *S. sonnei* sequences with publicly available sequences utilized in a previous global analysis (Holt *et al.*, 2012). An initial phylogeny was inferred with these data using RAxML under the GTRGAMMA model. To further assess the relationship of the Bhutanese isolates within Central Asia, the previous alignment was then subsampled to include only the 71 Bhutanese sequences along with 20 other closely related Central Asia sequences. Previously described mobile genetic elements (Holt *et al.*, 2012) and potential recombinant regions were removed from the alignment using Gubbins (Croucher *et al.*, 2014). Further removal of gaps, indeterminate bases, invariant sites and 17 Bhutanese sequences that were identical to others in the dataset resulted in a final alignment of 996 SNPs across 74 taxa. A maximum likelihood phylogeny was inferred from these

sequences using GARLI (Zwickl, 2006) with 1000 bootstrap replicates under the transversion (TVM) model of nucleotide substitution, which was determined to be the best-fit model using jModelTest (Posada, 2008).

In order to thoroughly examine the population and temporal structure of Bhutanese *S. sonnei* and the population from which they emerged, all sequences from Bhutan, including identical sequences removed in the previous phylogeny, were aligned with the six most closely related isolates from other countries and trimmed down to include only the 544 SNPs present in the alignment of 77 taxa. These data were then utilized for maximum likelihood reconstruction using PhyML under the TVM + Γ_4 substitution model (Guindon *et al.*, 2010). I proceeded to infer the evolutionary dynamics of Bhutanese *S. sonnei* using methods described in section 2.1.4. Multiple BEAST runs were conducted using the TVM or GTR nucleotide substitution model with constant, exponential growth, or Bayesian skyline demographic history priors, in combination with either a strict or a relaxed molecular clock (uncorrelated lognormal distribution) (Drummond and Rambaut, 2007). The most suitable model was TVM + Γ_4 substitution model with a relaxed lognormal molecular clock and a piecewise Bayesian skyline demographic model.

2.2.5 Gene content analysis

De novo assembly for each read set was constructed, and the accessory genome was explored as described in section 2.1.6. Multi-locus sequence typing (MLST) was performed for all *Shigella* isolates by profiling the seven MLST genes (*adk*, *fumC*, *gyrB*,

icd, *mdh*, *purA*, *recA*). Mutations in specific genes were manually assessed by retrieving their alignments from the annotations, and the effects of nonsynonymous mutations were predicted using the SIFT web server (http://sift.jcvi.org/) (Sim *et al.*, 2012).

2.3 Methodology specific to Chapter 4: The emergence and global spread of fluoroquinolone resistant *Shigella sonnei*

2.3.1 Organism collection and whole genome sequencing

This investigation regarding the international dissemination of FQ resistant *S. sonnei* was performed by combining WGS data from a collection of 411 *S. sonnei* isolates (Bhutan, n=71; Vietnam, n=24; Thailand, n=8; Cambodia, n=1; Ireland, n=20; Australia, n=85; France, n=97; England, n=91; global references, n=14). Susceptibility to ciprofloxacin was determined using a variety of methods, including disk diffusion, E-test, agar dilution, and broth microdilution (see Table 4.1). The breakpoint for ciprofloxacin resistance was inferred following guideline from the European Committee on Antimicrobial Susceptibility Testing (http://www.eucast.org/clinical_breakpoints/). For all *S. sonnei* isolates, genomic DNA was extracted using commercial kits and subjected to WGS on various Illumina platforms to produce paired-end short read sequences. These sequence data were then consolidated for analysis and interpretation.

2.3.2 Phylogenetic analysis

All reads were mapped to the *S. sonnei* Ss046 reference sequence (Accession number: NC_007384) using SMALT (version 0.7.4). High quality SNPs were identified as described in section 2.1.2. Preliminary phylogenetic reconstruction demonstrated that

among 411 examined sequences, only eleven clustered outside this CenAsiaIII clade. These eleven sequences, together with another two sequences with poor mapping quality and three outgroups to the CenAsiaIII clade, were further removed to create a final set of 395 sequences belonging to this clade. Recombinogenic elements were removed, along with previously determined mobile genetic elements, invariant sites, and columns with at least 2% indeterminate bases, to create a final alignment of 3,980 SNPs across 396 taxa (395 CenAsiaIII and one closely related out-group). Maximum-likelihood phylogenies were constructed in five separate runs using RAxML v8.1.3 under the GTRCAT model, with sufficient bootstrap replicates automatically determined by MRE bootstrap convergence (Stamatakis, 2014).

2.3.3 Hierarchical Bayesian clustering

Phylogeny-independent hierarchical Bayesian clustering (hierBAPS) was applied to the above alignment to identify population structure within the 395 CenAsiaIII isolates (Corander *et al.*, 2008; Cheng *et al.*, 2013). This was performed in ten independent analyses, with three layers of sub-clusters and maximum number of clusters ranging from 10 to 40. These analyses consistently identified two populations of identical membership in 8/10 instances, with all FQ resistant *S. sonnei* belonging to a single population (named Pop2).

2.3.4 Bayesian phylogenetic inference

Aiming to examine the temporal structure of the CenAsiaIII clade, I subsampled to include 144 sequences (72 from each hierBAPS determined populations, with a

maximum of five isolates per calendar year in each geographic region). This subsampling generated an alignment of 2,448 recombination-free SNPs, and a maximum-likelihood phylogeny was inferred. A TVM model of nucleotide substitution was determined to be the most appropriate for this alignment. The phylogeny showed a linear relationship between root-to-tip divergence and sampling date (R^2=0.845). BEAST v.1.8.3 was used to estimate the substitution rate and tMRCA of the CenAsiaIII clade (Drummond and Rambaut, 2007). Multiple BEAST runs were conducted on combinations of the TVM + Γ_4 substitution model, and strict, or relaxed lognormal clock model in conjunction with constant, or Bayesian skyline demographic models. The best model was the strict molecular clock with a piecewise Bayesian skyline demographic model.

I additionally repeated the same procedure for 141 selected Pop2 S. *sonnei* to accurately estimate the divergence time of FQ resistant clone. These analyses again indicated that the TVM + Γ_4 substitution model with a strict clock and Bayesian skyline demographic model was the most appropriate. I further selected a subset 134 FQ resistant *S. sonnei* to infer the population dynamic of the clone since its emergence to avoid artificial evidence of population expansion (Lapierre *et al.*, 2016). BEAST runs were performed in triplicate and conducted with the TVM + Γ_4 substitution model, a strict molecular clock with either constant, Bayesian skyline, skygrid, or skyride demographic model (Drummond *et al.*, 2005; Minin, Bloomquist and Suchard, 2008). The best-fit demographic model was determined to be Bayesian skyline.

2.3.5 The geographic spread of CenAsia III *Shigella sonnei*

The reference-based maximum likelihood phylogeny of 395 CenAsiaIII *S. sonnei* served as an input to reconstruct the ancestral geographical states of each isolate. The procedures were followed as described in section 2.1.7. Thirteen isolates from each location of South Asia, Africa, Europe, and combined Southeast Asia/East Asia were included for analysis. Samples from the Middle East and America were excluded.

2.3.6 Analysis on pseudogenes and gene loss

Pseudogenes can be formed by numerous mechanisms in *Shigella*, including gene deletion, gene disruption by insertion sequences (IS) or small indels, and substitutions leading to gain of stop codon or the loss of start codons. An FQ resistant CenAsiaIII isolate (sample 2012-02037) was selected for long-read nanopore sequencing to produce a high quality complete reference. The sequencing was performed by Dr. Kathryn Holt's lab in University of Melbourne, Australia. I compared the entire genome sequence of 2012-02037 to Ss046 reference using ACT to identify regions pertaining to genomic deletions or disrupted by IS elements, but discounting such events which occurred within IS elements. In total, I detected 16 candidate events (two genomic deletions, and 14 IS disruptions). The pan-genome of 395 CenAsiaIII *S. sonnei* sequences was constructed using Roary with default setup (Page *et al.*, 2015), and gene losses were queried in all 395 isolates to record their presence/absence within the genome sequences. For fourteen identified candidate genes with IS element disruption, the intactness of these were queried against all CenAsiaIII strains using ARIBA, setting the matching length cutoff to 95%.

To detect pseudogenes formed by substitutions and small indels, I extracted 393 CenAsiaIII variant calling format (vcf) files (two could not be retrieved) generated by mapping the reads to reference Ss046. Mutations were annotated by SnpEff v4.1b using the Ss046 reference and were retained if they passed the quality filtering step and were classified by SnpSift v4.1b as either, i) indels within genes, ii) loss of a start codon, or iii) gain of a stop codon (Cingolani *et al.*, 2012). Identified mutations within IS and repetitive elements were considered unreliable and removed from further analyses.

2.3.7 Substitution mutation analysis

For the 395 CenAsiaIII sequences, substitutions across the reference Ss046 were summarized and annotated using an in-house script. Substitutions present within pre-determined mobile, repetitive, or recombination elements were excluded, resulting in a set of 4,230 substitutions for examination. Due to the limited temporal sampling period, I used dN/dS ratio (non-synonymous to synonymous substitution rate) to detect potential signals of selection, adopted from a previously recommended approach, assuming the absence of recombination in the dataset (Lieberman *et al.*, 2011; Hedge and Wilson, 2016). Briefly, a phylogenetic tree was constructed using these 4,230 SNPs in RAxML v8.1.3, and PAML was used for ancestral state reconstruction of all SNPs (Yang, 2007). Substitutions were classified as intergenic, synonymous, or non-synonymous based on comparison of each SNP to its annotation and the reconstructed ancestral state. The dN/dS rate was adjusted for transition/transversion rate and codon frequencies under the NY98 model (Nielsen and Yang, 1998). Genes were determined to be “under selection”

if their adjusted dN/dS ratio were >1.5 or if they possessed none synonymous and at least five non-synonymous mutations.

2.3.8 Cytokine profiling

To evaluate whether infections with FQ resistant and susceptible *S. sonnei* elicit differential systemic inflammatory responses during early infection, I measured the concentrations of various inflammatory cytokines in plasma from a group of *Shigella* infected patients (44 FQ resistant and 16 FQ susceptible *S. sonnei* infections). These were inpatients recruited during a hospital-based surveillance of bacterial diarrheal diseases in Ho Chi Minh City, Vietnam (Duong *et al.*, 2016). A panel of nine pre-determined cytokines (TNF-α, CXCL8, IL-1β, IL-1Ra, IL-17A, IL-6, IL-10, IFN-γ, and IL-22) (R&D Systems) was measured in 50μL plasma following manufacturer's instructions (Luminex Molecular Diagnostics, Austin, TX, USA). Only measurements for CXCL8, IL-1Ra, IL-6, IL-22 were within the acceptable range of detection. Two isolates (FQ resistant) with consistently elevated concentrations for four these cytokines were removed as outliers. Cytokine measurements were log-transformed, and logistic regression was applied to identify if any of these cytokines exhibited a significant association with the *S. sonnei* FQ phenotype (42 FQ resistant vs. 16 FQ susceptible *S. sonnei*).

2.3.9 Experimental evolution

To examine how continuous exposure to quinolones may impinge on the evolution of *S. sonnei*, I selected a representative FQ resistant (CenAsiaIII clade) and susceptible (Global III clade) isolates. Both isolates, cipR-VN and cipS-VN, were isolated in 2015 as part of

a hospital-based diarrheal surveillance study conducted in Ho Chi Minh City, Vietnam (Duong *et al.*, 2016). Isolates were selected based on sequencing quality and the absence of plasmids encoding resistance to 3rd generation cephalosporins and macrolides (to discount potential interactions). The MIC to ciprofloxacin and nalidixic acid were reconfirmed by E-test and broth dilution. M9 minimal was used as culture medium for the experiment to limit over-proliferation. The M9 medium was supplemented with glucose as the sole carbon source (final conc. 0.4%), and niacin (final conc. 25 μg/mL) as *Shigella* is unable to synthesize this vitamin *de novo* (Ahmed, Sarker and Sack, 1988).

CipR-VN and cipS-VN were subjected to three longitudinal culture conditions in duplicate, i) M9 (glucose + niacin) medium only, ii) M9 (glucose + niacin) supplemented with 256 μg/mL of nalidixic acid, and iii) M9/B3 (glucose + niacin) supplemented with half the appropriate MIC of ciprofloxacin (0.125 μg/mL for cipS-VN and 4 μg/mL for cipR-VN). Cultures were incubated at 37ºC with continuous agitation for eight weeks; a sub-culture was performed every 24 hours by transferring 20μL into 5mL of fresh medium. This process resulted in approximately eight bacterial generations ($\log_2$[5,000/20]) per 24 hours. At the end of the experiment (D56), organisms were harvested from 2mL culture from each of the twelve conditions. Total genomic DNA was extracted from each sample using the Wizard Genomic DNA Extraction kit (Promega), and these were predicted to represent *S. sonnei* population diversity within each condition. Library preparation and deep sequencing by Illumina MiSeq (~100 coverage) were performed for all twelve samples to generate 2x250 bp paired-end reads. For each sample, reads were mapped to reference Ss046 by BWA-MEM, and high quality SNPs

were called by samtools and retained if they had a consensus quality >50, a mapping quality >30, and a read depth >3 (Li, 2013). Analyses were then performed independently for cipR-VN and cipS-VN, using existing sequence data as time point zero (D0). The Genome Analysis Toolkit (GATK) was employed to extract differing mutations between the paired sequences (D0 and D56), eliminating ones identified in all conditions (McKenna *et al.*, 2010). I additionally confirmed the fidelity of these mutations using Artemis and ACT. *De novo* assemblies were constructed for each isolate using SPAdes v3.9.0 (Bankevich *et al.*, 2012), and a pan-genome was constructed separately for cipR-VN and cipS-VN by Roary to identify genomic deletions.

2.4 Methodology specific to Chapter 5: Southeast Asian *Shigella* from a historical perspective

2.4.1 Sample collection and genome sequencing

In order to explore the evolutionary history of *Shigella* in Southeast Asia, I gathered a vast collection of *Shigella* isolates from continental Southeast Asian countries, including Thailand, Vietnam, Laos and Cambodia. *Shigella* isolates were retrieved from multiple collaborative institutions in the region, including: AFRIMS Thailand, N=1,239; Lao-Oxford-Mahosot Hospital-Wellcome Trust Research Unit (LOMWRU), N=43; Cambodia-Oxford Medical Research Unit (COMRU), N=19; Pasteur Institute in Cambodia, N=19; OUCRU Ho Chi Minh city, N=113. All these *Shigella* were isolated from patients, predominantly young children, of dysenteric diarrhoea in respective countries. For all these isolates, genomic DNA was extracted from pure bacterial culture by staff scientists of the corresponding institutions, and DNA was sent to the WTSI for

whole genome sequencing. In addition, these data was appended with sequences available from previous studies, including 119 *S. flexneri* and 252 *S. sonnei* (Holt *et al.*, 2013). In summary, the compiled collection includes 1,804 sequences (803 *S. sonnei* and 1,001 *S. flexneri*), which covers multiple regions in the area, as detailed in Table 2.1. These consist of *S. flexneri* of all reported phenotypic serotypes, except for serotype 6 as it is more closely related to *S. boydii*. The collection time ranges from 1980 to 2013, but this differs between geographies.

Region	**Timespan**	***S. flexneri***	***S. sonnei***	**Inclusive areas**
Bangkok and Central Thailand	1980-2013	280	227	Bangkok, Nonthaburi, Samut Sakhon, Nakhon Pathom
Northern Thailand	2005-2010	7	9	Chiang Rai, Mae Hong Son
Northeast Thailand	1980-2010	55	12	Nong Khai, Ubon Ratchathani, Surin, Nakhon Ratchasima
Western Thailand	1990-2002	186	82	Tak, Kanchanaburi, Ratchaburi
Southern Thailand	2004-2006	3	16	Trang
Cambodia	2004-2013	56	13	Phnom Penh, Siem Reap
Laos	2006-2012	34	9	NA
Northern Vietnam	1995-2004	147	56	Hanoi
Central Vietnam	2008-2011	28	49	Hue
South Central Vietnam	2002-2011	41	86	Khanh Hoa
Southern Vietnam	1991-2010	162	246	Ho Chi Minh city, Dong Thap, Tien Giang, Ben Tre

Table 2.1 Summary of isolates used in Southeast Asia *Shigella* evolution analysis

2.4.2 *Shigella flexneri* molecular serotyping scheme

S. flexneri is composed of multiple serotypes, and it is important to characterize the isolate's specific serotypes to quantify their burden and distribution. Serotyping has been conventionally determined by agglutination reactions between the isolate's LPS O-antigen and commercially available antisera (Denka Seiken, Japan). However, these could not cover all *S. flexneri* epitopes as novel serotypes arise from bacteriophage-mediated O-antigen modifications (section 1.7.1). I sought to exhaustively identify serotype determining genetic elements from *S. flexneri* whole genome sequences, adopting the methodology developed previously (Sun *et al.*, 2011; Gentle *et al.*, 2016). Genomic derived serotypes have shown a high concordance (>90%) with phenotypic ones, and the majority of discrepancy was due to disruptive genetic mutations (Chattaway *et al.*, 2017). In general, ARIBA was used to identify serotype specific determinants, including *gtrI*, *gtrII*, *gtrIV*, *gtrIc*, *gtrV*, *gtrX*, *oacA*, *oacB*, *oacD*, and *opt* (pSFXv_2). ARIBA outputs from each isolate were then compiled and curated. Genes that contained deletion, frameshift or nonsense mutations were deemed non-functional and counted as being absent from the isolate. Genomic serotype was defined based on specific combinations of these elements, as detailed in Table 2.2.

gtr (glucosylation)	***oac*** (O-acetylation)	***opt*** (phosphorylation)	**Genomic serotype**
gtrI			1a
gtrI	*oacA*		1b
gtrI + gtrIc			1c
gtrII	*oacD*		2a
gtrII + gtrX			2b
gtrII		*opt*	2av
gtrX	*oacA*		3a
	oacA		3b
gtrIV			4a
gtrIV	*oacA*		4b
gtrIV		*opt*	4av
gtrV			5a
gtrV + gtrX			5b
gtrX			X
gtrX		*opt*	Xv
			Y
		opt	Yv

Table 2.2 Geno-serotyping scheme for *Shigella flexneri*

2.4.3 Phylogenetic reconstruction

Phylogenetic reconstructions were performed independently for *S. flexneri* and *S. sonnei*, as mentioned in section 2.1.2 and 2.1.3. In order to provide phylogenetic context, I combined the Southeast Asian *S. flexneri* collection with more than 300 sequences from the species global evolution study (Connor *et al.*, 2015). Regions pertaining to mobile and recombinogenic elements were removed, so were alignment columns with at least 2% gaps or indeterminate bases. This created a 51,616 bp SNP alignment of 1,323 sequences, which was input into RAxML for phylogenetic reconstruction under a GTRCAT model with 100 bootstrap replicates. Phylogenetic groups (lineages) were defined as previously proposed. To identify the most appropriate root for the *S. flexneri* species tree, I constructed a maximum likelihood phylogeny based on the core genome SNP alignment of 111 representative sequences. These include ten random samples from each defined lineages (1, 2.1, 2.2, 3Xv, 3.1, 3.2, 4, 5, 6, 7, 8) and a *S. flexneri* serotype 6 isolate (used as an outgroup). The core genome alignment was built using Roary, which utilized the annotations of *de novo* assemblies (Page *et al.*, 2015). The resulted phylogeny indicated that the common ancestor of Lineages 7 and 8 serve as the most appropriate root for *S. flexneri* species phylogeny. Maximum likelihood phylogenies were generated independently in similar manner for lineages containing substantial number of Southeast Asian sequences, including Lineages 1, 2, 3Xv, 3.1, 3.2, 4, 7, and 8.

For *S. sonnei*, I combined my collection with 190 representative sequences from the species global evolution study (Holt *et al.*, 2012). I followed the aforementioned procedures to produce a 14,241 bp SNP alignment of 993 sequences. This was in turn

used for phylogenetic reconstruction in RAxML under a GTRCAT model, with 100 bootstrap replicates. Phylogenetic nomenclatures followed ones established previously in the global study, and I constructed individual maximum likelihood phylogenies for the following groups: Lineage 1, Lineage 3 clades earlySEA, Global 3.1 and Global 3.2.

2.4.4 Bayesian phylogenetic inference

I conducted temporal analysis and Bayesian phylogenetic inferences separately for each major *Shigella* lineages. At most five isolates per year per region were included for each lineage to limit sampling bias. Table 2.3 provides detailed information on the input for each analysis. BEAST v.1.8.3 was used to estimate the substitution rate and tMRCA of each lineage (Drummond and Rambaut, 2007). Multiple BEAST runs were conducted on combinations of the appropriate substitution model, and a strict, random local clock or relaxed lognormal clock model in conjunction with a constant, or Bayesian skyline demographic model.

Species	**Lineage**	**No. of sequences**	**Alignment length (bp)**	**Substitution model**
flexneri	1	78	3,495	TVM + Γ_4
flexneri	2.1	32	471	K3P + Γ_4
flexneri	2.2	68	1,986	TVM + Γ_4
flexneri	3Xv	32	1,741	TVM + Γ_4
flexneri	3.1	76	1,586	TVM + Γ_4
flexneri	3.2	117	3,842	TVM + Γ_4
flexneri	7	49	1,844	TVM + Γ_4
flexneri	8	55	1,015	TVM + Γ_4
sonnei	Lin1	45	2,860	TVM + Γ_4
sonnei	earlySEA	51	675	TVM + Γ_4
sonnei	Global 3.1	151	1,633	TVM + Γ_4
sonnei	Global 3.2	115	1,779	TVM + Γ_4

Table 2.3 Input alignment for each *Shigella* lineage used for Bayesian phylogenetic inferences

2.4.5 Inter-regional dissemination of *Shigella* lineages

I explored the migration pattern of major *Shigella* lineages in Southeast Asia, including *S. flexneri* Lineages 3.1 and 3.2, and *S. sonnei* Lineages Global 3.1 and 3.2. The ancestral geographical state of each isolate was deduced by stochastic mapping, using the respective lineage's maximum likelihood phylogeny. The procedures were as described in section 2.1.7. Table 2.4 details the input information for each analysis.

Lineage	**No. of isolates in input tree**	**No. of isolates per region**	**Inclusive geographical region**	**No. of subsampled trees**
S. flexneri 3.1	237	11	BKK, CAM, HUE, KH, LAO, NET, NVN, SVN	500
S. flexneri 3.2	281	14	BKK, LAO, NET, NVN, SCVN, WT	500
S. sonnei Global 3.1	391	12	BKK, CAM, HUE, KH, NVN, SVN	500
S. sonnei Global 3.2	176	18	BKK, SVN, WT	500

Table 2.4 Input information for stochastic mapping of geographical states, separately for each major *Shigella* lineage. BKK: Bangkok and Central Thailand; NET: Northeast Thailand; WT: Western Thailand; CAM: Cambodia; LAO: Laos; NVN: Northern Vietnam; HUE: Hue; KH: Khanh Hoa; SVN: Southern Vietnam; SCVN: Southern and Central Vietnam (combined Southern Vietnam, Hue, and Khanh Hoa).

2.5 Methodology specific to Chapter 6: Assessing gut microbiota perturbations during the early phase of infectious diarrhoea in Vietnamese children

2.5.1 Study design, sample collection and microbiological procedures

Samples for this study originated from a previously described prospective observational study of paediatric diarrhoea conducted at three major hospitals in Ho Chi Minh City (HCMC), Vietnam: Children's Hospital 1 (CH1), Children's Hospital 2 (CH2), and the Hospital for Tropical Diseases (HTD) (Thompson *et al.*, 2015). The Oxford Tropical Research Ethics Committee (OxTREC) and the Scientific and Ethics Committee of CH1, CH2, and HTD provided ethical approval for the primary and this subsequent analysis. The general inclusion criteria for both diarrhoeal cases and non-diarrhoeal controls were children aged under five years, residing within HCMC and reporting no antimicrobial usage within three days prior to hospital admission. Children admitted to the study sites with diarrhoea (defined as three or more loose stools or at least one bloody loose stool within a 24 hour period) were included as cases. Children who presented for health checks, nutritional, or gastrointestinal issues but reported no diarrhoeal or respiratory illnesses within seven days of admission were recruited as controls. For each enrolee, clinical data regarding the symptoms and duration of diarrhoea was obtained from a case report form completed by study clinicians, while demographic, feeding behaviour and socioeconomic details were provided through a confidential questionnaire. Weight-for-age Z (WAZ) score was used to evaluate the nutritional status of all enrolled children based on WHO standards (World Health Organization, 2006).

A faecal sample was collected from each participant, both cases and controls, prior to any prescribed antimicrobial treatment. Both case and control faecal samples were subjected to standard microbiological culturing and biochemical testing to identify common diarrhoeal bacteria, including *Shigella*, *Salmonella*, *Campylobacter*, *Plesiomonas*, *Yersinia*, and *Aeromonas*, as previously described (My *et al.*, 2013) (https://wwwnc.cdc.gov/EID/article/19/6/11-1862-Techapp1.pdf). For the detection of Rotavirus and Norovirus, reverse transcriptase polymerase chain reaction (RT-PCR) was performed on total viral RNA extracted from all specimens (My *et al.*, 2011). Further detection of *Giardia lamblia*, *Entamoeba histolytica*, and *Cryptosporidium* cysts were performed by microscopy on fresh faecal samples diluted in phosphate buffered saline. All described laboratory work were conducted at OUCRU by staff scientists. Stool samples were stored at -80°C until being subjected to DNA extraction up to six months after initial storage.

2.5.2 DNA extraction and sequencing

The prospective study from which the faecal specimens originated successfully recruited 1,419 cases and 609 controls. A subset of 200 faecal samples was selected for this investigation; these included 55 randomly selected controls and 145 cases from multiple confirmed aetiologies. For diarrhoeal cases, random subsamples from each identified aetiology were included to represent the epidemiology reported in a previous prospective study (Thompson *et al.*, 2015), with emphasis on *Shigella* as it was the major cause of dysentery. Total DNA was extracted from these faecal samples by a staff scientist at OUCRU, using a phenol-chloroform extraction method. Briefly, 200µl of the faecal

sample was suspended in a solution containing 500µl of DNA extraction buffer, 50µl NaCl 200mM, 50µl SDS 10%, 500µl of phenol:chloroform:isoamyl (25:24:1), and 0.7 g of 0.5-mm-diameter zirconia/silica beads. Cells were lysed by mechanical disruption with a bead beater for one minute and subjected to three rounds of phenol:chloroform extraction. DNA was resuspended in TE buffer supplemented with RNase.

All DNA samples were shipped on dry ice to the Genome Institute of Singapore (GIS) for 16S rRNA sequencing using the GIS Efficient Rapid Microbial Sequencing (GERMS) platform. In order to obtain high resolution taxonomic identification up to the species level, all samples were PCR amplified using a previously optimized primer set (338F-1061R), which produces >700 bp amplicons covering the four variable regions (V3-V6) of the 16S rRNA region and can be assembled *in silico* to retrieve > 92% of the sequences in the Greengenes database (Ong *et al.*, 2013). Long PCR fragments were cleaned by GIS staff, using 1X AMPURE beads and randomly fragmented with Covaris (model LE220) shearing to ~200bp. Library preparation was performed using the GeneRead DNA Library I Core Kit (Qiagen, Germany) according to the manufacturer's instructions, and the prepared DNA was subjected to sequencing on an Illumina HiSeq2500 platform to produce 75 bp paired-end reads.

2.5.3 EMIRGE assembly of 16S rRNA amplicons and OTU clustering

EMIRGE (Expectation Maximization Iterative Reconstruction of Genes from the Environment) is a database dependent assembler built on an iterative expectation-maximization (EM) algorithm, which is used to reconstruct full length 16S rRNA small

subunit gene sequences (by simultaneously mapping and clustering) and estimate their relative abundance (Miller *et al.*, 2011). The SILVA small subunit (SSU) rRNA database version 119 was filtered to remove potential large subunit (LSU) rRNA sequences, and closely related sequences were clustered at 97% identity by USEARCH; this was achieved by using PhyloFlash v2.0 (https://github.com/HRGV/phyloFlash). The resulting database was used as the reference template for EMIRGE. To limit the computational effort related to utilizing the complete dataset, one million paired-end reads were randomly subsampled without replacement from each sample library using seqtk (https://github.com/lh3/seqtk). Reads were trimmed using Sickle to remove those with quality <30 and length <60 (Joshi and Fass, 2011). I inputted the trimmed reads from each subsampled library into the amplicon-optimized version of EMIRGE, with 40 iterations and a 97% joining threshold. The EMIRGE output, a set of assembled and clustered sequences, for each sample indicates its representative OTUs (Operational taxa units) with their estimated abundances. Assembled sequences with sample-wise normalized relative abundances less than 0.01% were removed from further analysis.

A pseudo-count for each sequence per subsampled library was calculated by scaling the number of successfully mapped reads to the EMIRGE estimated relative abundance of each sequence. All filtered sequences and a count table detailing their respective abundance were pooled and imported into the 16S rRNA processing platform mothur v.1.36.0 (Schloss *et al.*, 2009). In order to minimize the length differences in assembled sequences and to facilitate more accurate OTU clustering, sequences were aligned to a trimmed version of the SILVA reference database to include only the amplified region

(338F-1061R). Gap-only columns were filtered from this alignment, and sequences with a maximum of seven (~1%) ambiguous sites were retained for downstream analyses. Sequences were dereplicated, and ambiguous sites were replaced randomly with one of the four nucleotides (ATCG). These unique sequences and their respective counts served as input for the UPARSE clustering algorithm, which clustered all pooled sequences at a 97% similarity threshold (Edgar, 2013). Chimeric sequences were stringently predicted and removed by UCHIME v6.0, set at both *de novo* mode and against the ChimeraSlayer informed reference database (e.g. 'gold' database). A total of 7,479 OTUs were reconstructed for 199 of the successfully sequenced samples. An alignment consisting of the most abundant representative sequences from each OTU was used to construct a phylogenetic tree using FastTree 2 under default parameters (Price, Dehal and Arkin, 2010). Taxonomic assignments of OTU representatives up to the genus level were performed using the mothur implemented Ribosomal Database Project (RDP) classifier, with a minimum support threshold of 80% (Wang *et al.*, 2007).

2.5.4 Data analysis

All analyses were conducted in R (R Core Team, 2016) using multiple packages, including 'phyloseq', 'cluster', 'randomForestSRC', 'ggplot2', 'nnet', 'lmtest', 'vegan', 'DESeq2', 'SpiecEasi' and other packages (Venables and Ripley, 2002; Zeileis and Hothorn, 2002; Wickham, 2009; McMurdie and Holmes, 2013; Love, Huber and Anders, 2014; Kurtz *et al.*, 2015; Ishwaran and Kogalur, 2016; Maechler *et al.*, 2016; Oksanen *et al.*, 2016). An OTU count table, taxonomy classification table, related clinical and demographic data and the OTU phylogenetic tree were imported and analysed as a

'phyloseq' object, allowing a unified and interactive analysis approach (McMurdie and Holmes, 2013).

2.5.5 Clustering into Community State Types (CSTs)

In order to gain overall insight into the microbial compositional structure of both controls and diarrhoeal cases, I applied a previously described clustering approach (DiGiulio *et al.*, 2015). To reduce sparsity as well as account for the divergence and the functional similarity shared between members of the same genus, which is frequently reported for human gut microbiota, the OTU count table was agglomerated at the unique genus level for clustering as recommended (Koren *et al.*, 2013). Recent benchmark studies have proposed that the weighted Unifrac distance produces desirable accuracy and power in exploring beta-diversity (McMurdie and Holmes, 2014; Thorsen *et al.*, 2016; Weiss *et al.*, 2017). A weighted Unifrac dissimilarity matrix between all samples was calculated using the collapsed genera's relative abundances and their phylogenetic relatedness (Lozupone *et al.*, 2007). The partitioning around meloids (pam) algorithm was applied to this matrix to cluster all samples into distinct CSTs, with the optimal number of CSTs (k=4) determined using gap statistics, a goodness of clustering measure (Tibshirani, Walther and Hastie, 2001). Further assessments such as prediction strength, average silhouette width (asw) were employed to validate this optimal number of clusters, as recommended previously (Koren *et al.*, 2013). Microbial CST grouping has been applied extensively to profile the human vaginal microbiome and has produced consistent findings (Ravel *et al.*, 2011; DiGiulio *et al.*, 2015; MacIntyre *et al.*, 2015). To assess the performance of this clustering approach and identify the most important genera

contributing to the separation of these CSTs, the random forest classification algorithm was applied to the genera relative abundance table, using the samples' CST memberships as the response variable (Breiman, 2001). A single tree assembly (5,000 trees) was used for all 199 samples of four CSTs, and other parameters were set as default according to the 'rfsrc' function in package 'randomforestSRC' (Ishwaran and Kogalur, 2016).

2.5.6 Statistical analysis and regression modelling

In order to examine the relationship between the microbiome structures and associated explanatory variables for 195 samples with complete metadata, I applied constrained analysis of principal coordinates (CAP) on the calculated weighted Unifrac dissimilarity matrix and a set of demographic variables (age, sex, WAZ score, feeding pattern, income, rural residence), as well as diarrhoeal status. This was performed using the 'capscale' function in the 'vegan' package (Egendre and Anderson, 1999; Oksanen *et al.*, 2016). Significant variables were identified and included in the final model based on Akaike information criterion (AIC) in a stepwise model selection approach. The alpha diversity of these 195 samples was estimated by Shannon diversity index. Analysis of variance (ANOVA) with post-hoc Tukey test and Bonferroni correction for multiple comparisons was used to compare control and diarrhoeal alpha-diversity among and within each CST. Multinomial logistic regression modelling was applied using the 'multinom' function in 'nnet' package to evaluate the association of the various aforementioned demographic factors and clinical features (vomiting, dysentery, and infection type) to the CST membership of diarrhoeal cases, with the *Bacteroides* rich CST (CST2) serving as the reference group (Venables and Ripley, 2002). These predictors were included on the

basis of limited missing data and capable of being assessed by clinicians upon patient's admission. Five samples with WAZ scores >3 or <-3, and one CST1 sample with age exceeding two standard deviations were considered as outliers and removed, resulting in 136 diarrhoeal samples with full metadata being subjected to regression modelling.

2.5.7 Evaluating differential abundance

DESeq2 was utilized to normalize the 7,479-OTUs table and to detect OTUs that show significantly differential relative abundance between two assessed groups per comparison (diarrhoea against controls, bacterial infection against viral infection, and dysentery against non-dysentery), as recommended in recent benchmark studies (Love, Huber and Anders, 2014; McMurdie and Holmes, 2014; Thorsen *et al.*, 2016). To identify consistent trends observed across all CSTs (diarrhoea against controls, bacterial infection against viral infection) or infection types (dysentery against non-dysentery), I applied the likelihood ratio test (LRT) approach with the reduced models incorporating only either CST or infection type as coefficients. OTUs with adjusted p values <0.05, estimated fold change >4 or <1/4, and estimated base mean >30 were considered significantly differentially abundant between two examined classes. Taxonomic identification to the species level was performed by manually comparing the reconstructed 16S rRNA sequences of OTUs to either the Human Oral Microbial Database (HOMD; http://www.homd.org/) or the Human Microbiome Project (HMP; http://www.hmpdacc.org/resources/blast.php) by BLAST. OTU was assigned a species nomenclature if its sequence is of at least 97% similarity to that in the database. Using these approaches, *Shigella* sequences are indistinguishable from *Escherichia.*

2.5.8 Interaction network construction

A correlation network was constructed for all 199 control and diarrhoeal samples to characterize the potential interactions between most representative 92 OTUs, defined as OTUs detected in at least 10 samples. This filtering step did not substantially affect the representativeness of the dataset, with the median sample retainment rate of 93% (IQR: 87% - 97%). The network was constructed using the SparCC wrapper in the package 'SpiecEasi' (Friedman and Alm, 2012; Kurtz *et al.*, 2015). The statistical significance for each interaction was assessed by 100 bootstrap iterations, with *p* values adjusted for multiple comparison correction. To avoid spurious correlations, only those with adjusted *p* values no greater than 0.05 and absolute magnitude equal to or above 0.25 were considered as significant correlations and represented in the final plots.

3 The introduction and establishment of fluoroquinolone resistant *Shigella sonnei* into Bhutan

All the work described in Chapter 3 has been published during my DPhil, under an article in the journal Microbial Genomics named:

Chung The, H., Rabaa, M., Pham Thanh, D., Ruekit, S., Wangchuk, S., Dorji, T., Tshering, K. P., Nguyen, T. N. T., Vinh, P., Thanh, T., Minh, C., Turner, P., Sar, P., Thwaites, G., Holt, K. E., Thomson, N. R., Bodhidatta, L., Mason, C. J. and Baker, S. (2015) 'Introduction and establishment of fluoroquinolone resistant *Shigella sonnei* into Bhutan', *Microbial Genomics*, 1(6), pp. 1–11. doi: 10.1099/mgen.0.000042

3.1 Abstract

Shigella sonnei is a major contributor to the global burden of diarrhoeal diseases, generally associated with dysenteric diarrhoea in developed countries but also emerging in developing countries. The reason for the recent success of *S. sonnei* is unknown but is likely catalysed by its ability to acquire resistance against multiple antimicrobials. Between 2011 and 2013, *S. sonnei* exhibiting resistance to fluoroquinolones (FQ), the first-line treatment recommended for shigellosis, emerged in Bhutan. Aiming to reconstruct the introduction and establishment of FQ resistant *S. sonnei* populations in Bhutan, I performed whole genome sequencing on 71 *S. sonnei* isolated in Bhutan between 2011 and 2013. I found that these strains represented an expansion of a clade within the previously described lineage III, found specifically in central Asia. Temporal

phylogenetic reconstruction demonstrated that all of the sequenced Bhutanese *S. sonnei* diverged from a single ancestor that was introduced into Bhutan around 2006. The analysis additionally predicted that FQ resistance, conferred by mutations in *gyrA* and *parC*, arose prior to the introduction of the founder strain into Bhutan. Once established in Bhutan, these *S. sonnei* had access to a broad gene pool, as indicated by the acquisition of Extended Spectrum Beta-lactamase encoding plasmids and genes encoding type IV pili. The data presented here outline a model for the introduction and maintenance of FQ resistant *S. sonnei* in a new setting. Given the current circulation of FQ resistant *S. sonnei* in Asia, I speculate that this pattern of introduction is being recapitulated across the region and beyond.

3.2 Introduction

Recent advances in WGS and analytical methodologies have enhanced our understanding of the epidemiology, international spread and antimicrobial susceptibility trends of *S. sonnei* (Holt et al. 2012; Holt et al. 2013). We now know that *S. sonnei* is formed of three main lineages (I, II, III), which are all related to a common ancestor that arose in Europe approximately 400 years ago (Holt *et al.*, 2012). Lineage III, which is MDR, now dominates globally (Holt 2012). It was previously shown that once introduced into a new geographic region, into an apparently naïve human population, *S. sonnei* can establish new isolated clonal populations which evolve locally (Holt *et al.*, 2013). However, such detailed investigations remain insufficiently documented in other Asian countries, where shigellosis is endemic. Recently, antimicrobial resistance is considered a key driver of the recent *S. sonnei* global expansions (Pu *et al.*, 2015).

Bhutan is a lower middle-income country in South Asia, sandwiched between China and the northwest Indian states of Assam and West Bengal. In close proximity to Nepal (<70 km of Indian territory separates the two countries), Bhutan is a Himalayan nation with an estimated population of 750,000 people, of which the majority (69%) live in remote rural locations (Office of the Census Commissioner – Royal Government of Bhutan, 2006). Like other low-income countries in the region, diarrhoeal disease is a major public health problem in Bhutan, and both *S. flexneri* and *S. sonnei* are endemic (Marfin *et al.*, 1994; Ruekit *et al.*, 2014). Between March 2011 and October 2013, a diarrhoeal disease surveillance study was conducted in the Bhutanese capital, Thimphu. During this study, there was an increased number of *S. sonnei* isolated between June-July. 29 of these *S. sonnei* isolates, which were isolated during a presumed outbreak in Thimphu in June and July 2011, were identical by PFGE pattern and antimicrobial susceptibility (Ruekit et al. 2014). All these *S. sonnei* isolates were resistant to ciprofloxacin (MIC ≥3μg/μl), which was infrequently observed in strains circulating globally during the same period (Ruekit et al. 2014). This study suggested that these *S. sonnei* isolates were closely related and may have been a recent introduction into Bhutan.

To test the hypothesis that these FQ resistant *S. sonnei* had recently been introduced to Bhutan and to further characterize these Bhutanese isolates in the context of global *S. sonnei*, I performed WGS on 71 *S. sonnei* isolated in Bhutan between 2011 and 2013.

3.3 Results

3.3.1 Bhutanese *Shigella sonnei* lay within lineage III (Central Asia clade)

All *S. sonnei* (n=74) isolated during this study were subjected to WGS; 71 isolates produced sufficient sequence for analysis. To contextualize the Bhutanese *S. sonnei* within the known global *S. sonnei* population structure, I constructed a phylogeny that included the Bhutanese sequences and those from a previous global *S. sonnei* sequencing project (Holt et al. 2012); within this phylogeny, all Bhutanese *S. sonnei* fell into a single clade within lineage III (Central Asia Clade - CenAsiaIII), which is a sister clade to the intercontinentally expanding Global III Clade (Fig 3.1A). To investigate the phylogenetic relationships within the CenAsiaIII clade, I combined 74 of the non-duplicate whole genome sequences into one alignment, including 54 from Bhutan and 20 selected from other countries. A maximum likelihood phylogeny constructed from this alignment showed a close clonal relationship between the Bhutanese *S. sonnei*. Showing a similarly close relationship to the Bhutanese isolates, and interspersed within this clade, were also two isolates from Thailand and Cambodia, likely representing importations following regional travel (Fig 3.1B). The alignment was also examined using Gubbins to detect the presence of potential recombination regions within these genomes. The result showed that recombination is infrequently observed in this clonal *S. sonnei* population (Fig 3.2).

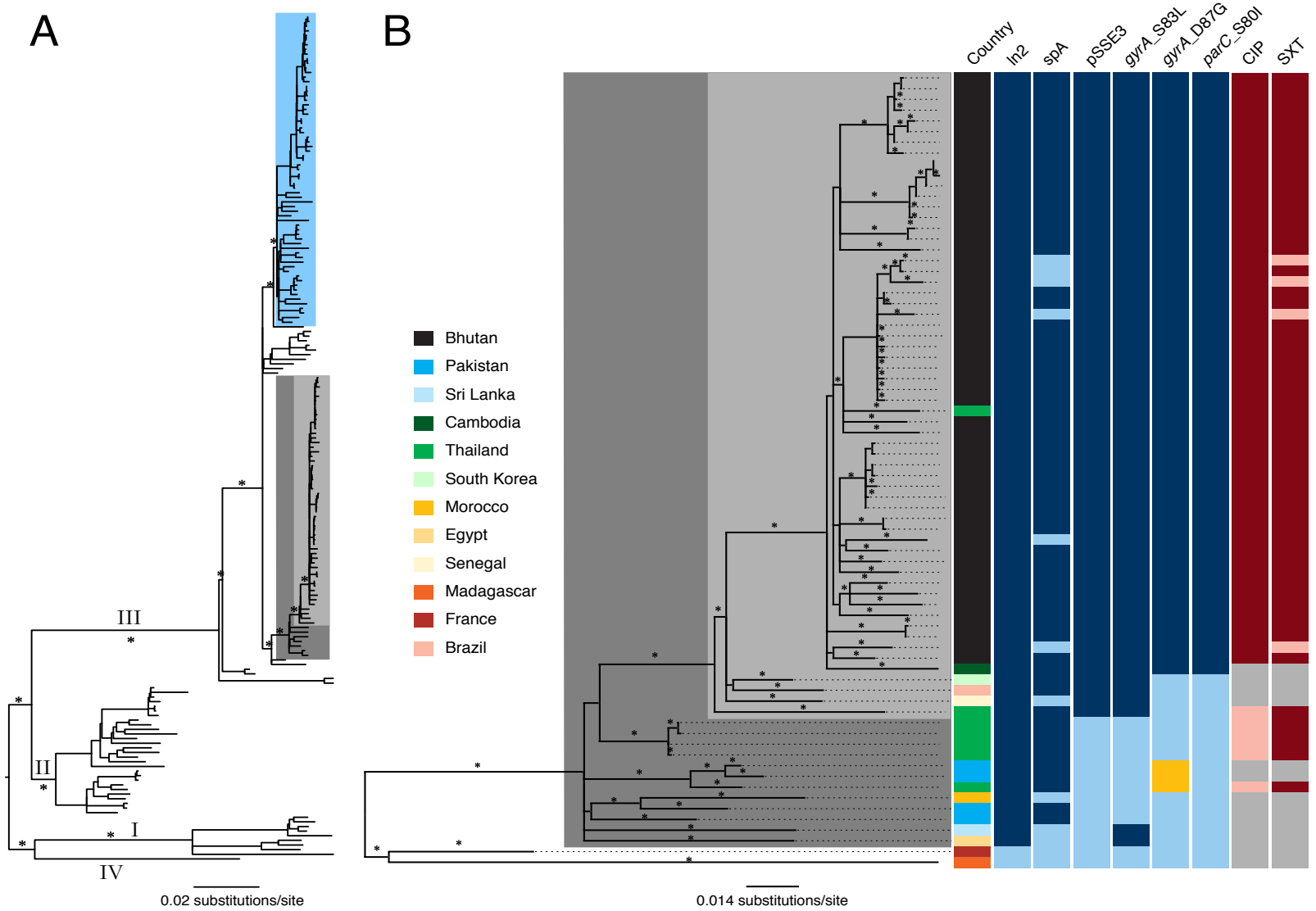

A
B
III
II
I
IV
0.02 substitutions/site
0.014 substitutions/site
Bhutan
Pakistan
Sri Lanka
Cambodia
Thailand
South Korea
Morocco
Egypt
Senegal
Madagascar
France
Brazil
Country
In2
spA
pSSE3
gyrA_S83L
gyrA_D87G
parC_S80I
CIP
SXT

Figure 3.1. The phylogenetic structure of Bhutanese *Shigella sonnei* in the context of the global phylogeny and CenAsiaIII clade

A) Midpoint rooted maximum likelihood phylogenetic tree of 183 *S. sonnei* strains (135 from global collection and 48 from Bhutan) constructed using 5,393 SNPs; asterisks indicate bootstrap support values ≥98% on major branches. Numbers above major branches represent ones leading to major lineages (I, II, III, IV). The light blue box highlights the Global III clade; the dark grey box overlaid on the tree highlights strains belonging to the CenAsiaIII clade; the smaller, light grey box highlights the primarily Bhutanese CenAsiaIII clade. B) Magnified view of the maximum likelihood phylogenetic tree of the CenAsiaIII clade, including 74 *S. sonnei* strains (54 from Bhutan and 20 others for phylogenetic context), constructed using 996 SNPs; asterisks indicate bootstrap support values ≥80%. The tree is midpoint rooted for purpose of clarity. Columns to the right of the phylogenetic tree show isolate metadata including: country of isolation (colours indicated in the key); the presence (dark blue) or absence (light blue) of specific genetic elements and mutations (In2, spA, pSSE3, *gyrA*-S83L, *gyrA*-D87G, *parC*-S80I), or the presence of an alternative mutation (yellow), *gyrA*-D87Y; and resistance profiles against ciprofloxacin (CIP) and trimethoprim/ sulfamethoxazole (SXT), where resistance is indicated in dark red, susceptibility in light red, and missing data in grey.

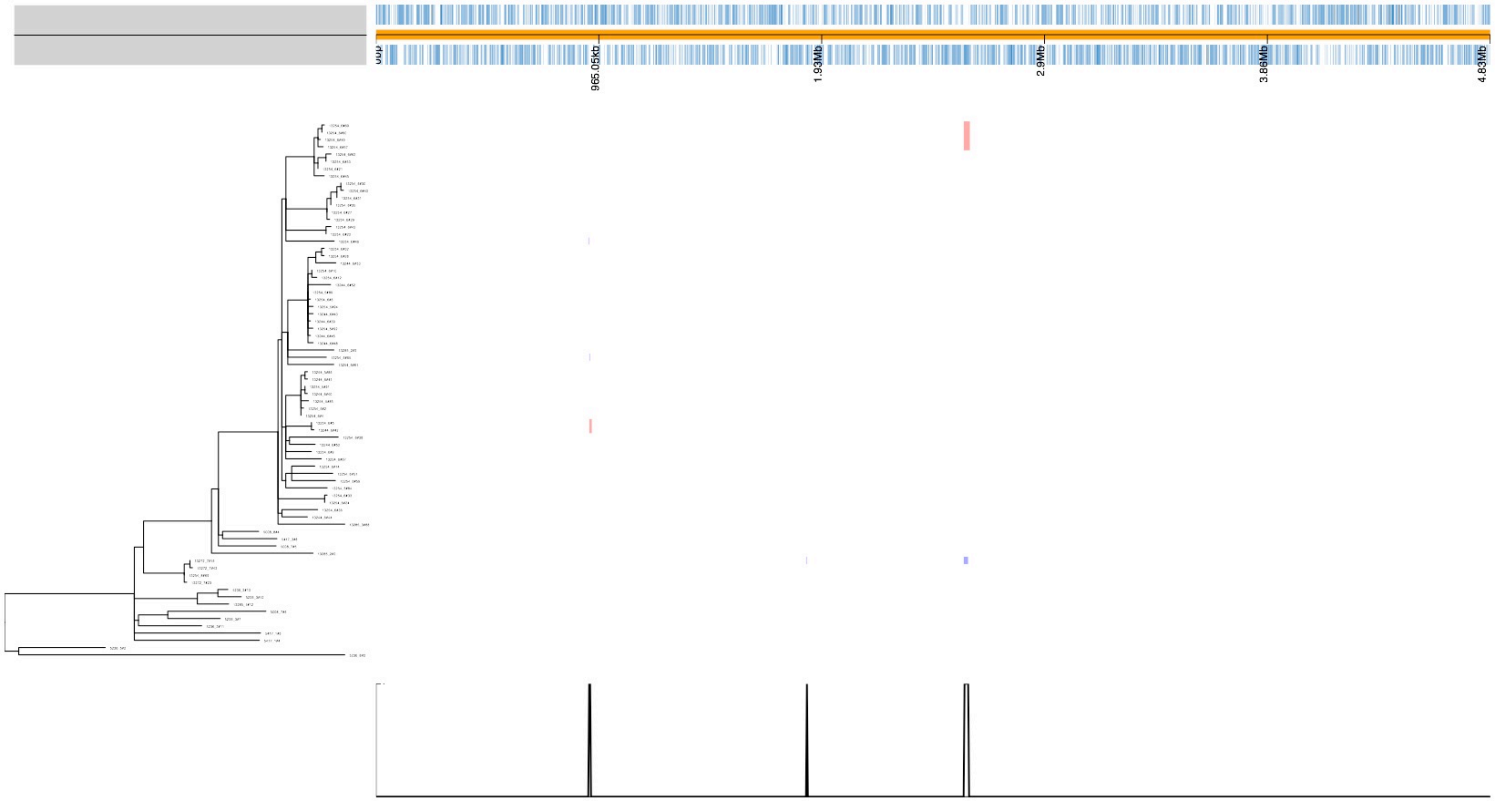


Figure 3.2 Recombination features within the genomes of CenAsiaIII *Shigella sonnei*.

The phylogeny on the left is identical to one described in Figure 3.1B. The top right bar displays the genomic coordinates of *S. sonnei* reference Ss046 (~4.3 Mbp), with coding sequences coloured in blue. Presumed recombination regions are detected within the examined genomes using Gubbins, which are highlighted in light-blue (occurring in single taxon) or red (shared between multiple monophyletic taxa).

3.3.2 Temporal phylogenetic reconstruction of *Shigella sonnei* in Bhutan

Since there was such a strong temporal signature for the number of observed mutations in the global *S. sonnei* population (Holt *et al.*, 2013), I explored the temporal structure of the *S. sonnei* isolated in Bhutan using Bayesian phylogenetic methods (section 2.2.4). To do this, I jointly estimated the nucleotide substitution rate and MRCA. The primarily Bhutanese lineage of CenAsiaIII evolved with a rate of 7.34×10^{-7} substitutions per site per year (95% highest posterior density (HPD): 5.87×10^{-7} to 8.83×10^{-7}), which is comparable to the previously calculated substitution rate of *S. sonnei* lineage III (8.5×10^{-7}; 95% HPD 7.6×10^{-7} to 9.5×10^{-7}) (Holt *et al.*, 2013). Using these data, I estimated that the MRCA of the Bhutanese *S. sonnei* existed around 2006 (95% HPD 2005 – 2007), with the majority diverging from a hypothetical MRCA around 2007 (95% HPD 2005 – 2008) (Fig 3.3). The results of this analysis were further supported by the linear relationship between root-to-tip branch length and date of isolation (R^2=0.8412). Since its projected introduction in 2006, this *S. sonnei* lineage diverged into at least four sub-lineages (labelled A, B, C and D in Fig 3.3), which have been transmitted endemically within the Bhutanese capital, Thimphu: isolates taken from cases in June/July 2011 (25/71 isolates) were found to belong to two subpopulations (A and C). Sublineage B, which was not represented in June/July 2011, dominated in the subsequent years.

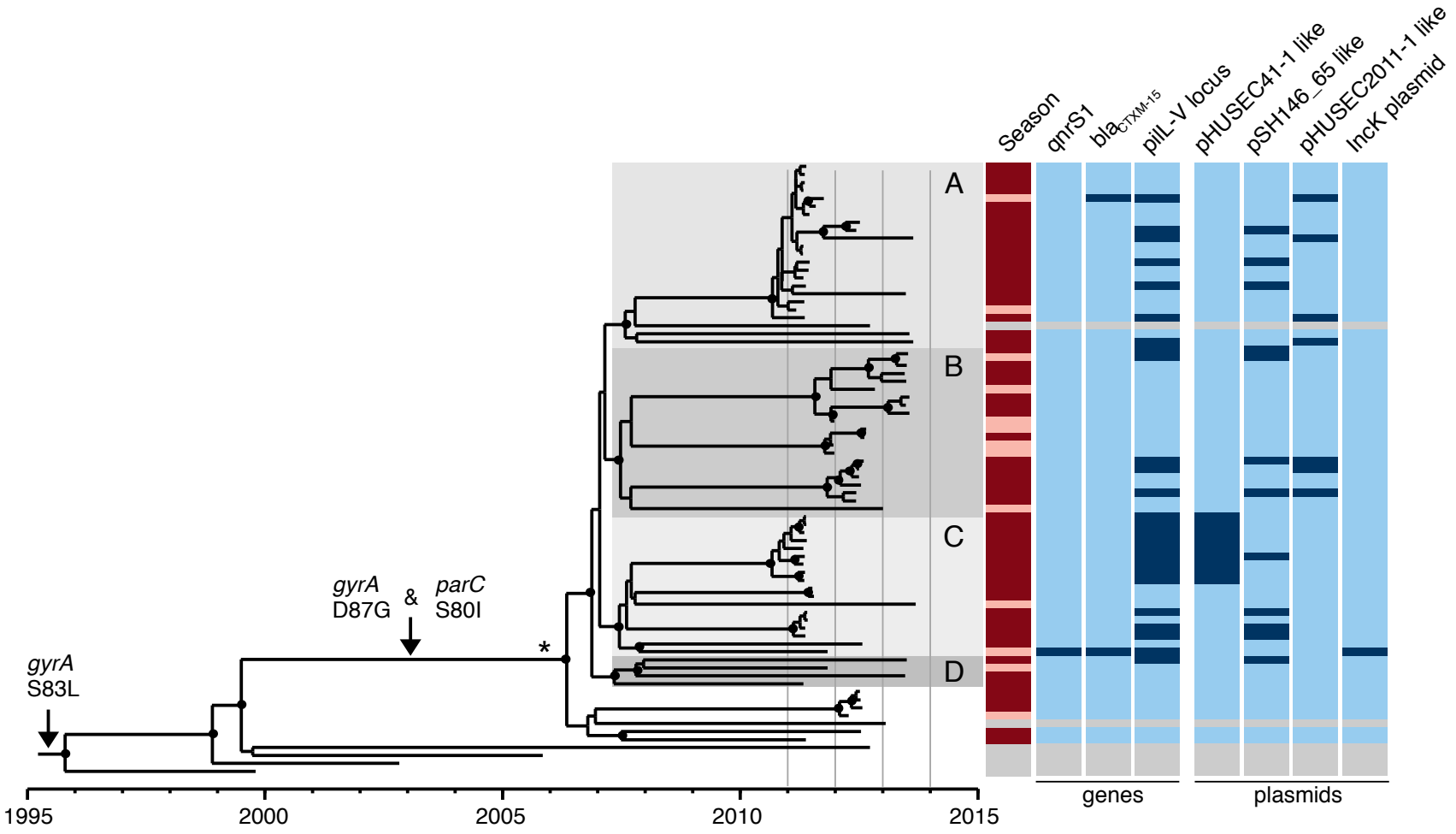

Season
qnrS1
blaCTXM-15
pilL-V locus
pHUSEC41-1 like
pSH146_65 like
pHUSEC2011-1 like
IncK plasmid
A
B
C
D
gyrA
S83L
gyrA
D87G
&
parC
S80I
*
1995
2000
2005
2010
2015
genes
plasmids

Figure 3.3 Temporal phylogenetic reconstruction of Bhutanese *Shigella sonnei* between 2011 and 2013

Image shows a maximum clade credibility (MCC) phylogenetic reconstruction of *S. sonnei* isolated primarily in Bhutan over the study period. Distinct subpopulations of Bhutanese *S. sonnei* are highlighted and indicated by letters. Black circles indicate posterior probability support ≥80% on internal nodes. The asterisk indicates the branch leading to the Bhutanese *S. sonnei* clade, which represents 22 lineage-defining SNPs. Arrows denote branches characterized by select substitution events. Columns to the right of the phylogeny show metadata including: season (dark red, monsoon season: June - September; light red, other); and presence (dark blue) or absence (light blue) of mobile genetic elements (*qnrS1*, $bla_{\text{CTXM-15}}$, pHUSEC41-1 like plasmid, pSH146_65 like plasmid, pHUSEC2011-1 like plasmid, incK plasmid). Grey bars in metadata columns indicate sequences from *S. sonnei* isolated outside of Bhutan.

3.3.3 Mutations specific to Bhutanese *Shigella sonnei*

To identify potential genetic mechanisms associated with diversification of the Bhutanese *S. sonnei,* I compared their content of unique SNPs in comparison to the three most closely related outgroup isolates, taken from global isolates belonging to the same global lineage but collected outside of Bhutan (Fig 3.3). The Bhutanese isolates were distinguished from these by the presence of 22 lineage-specific consensus SNPs. Of these 22 SNPs, five fell in intergenic regions and 17 mapped within predicted coding sequences; 11 represented nonsynonymous and six were synonymous mutations. The genes bearing nonsynonymous mutations are summarized in Table 3.1 together with their predicted impact on function. These include mutations in *gyrA* and *parC* associated with antimicrobial resistance, as well as those located in *ycgB*, *mppA*, *hisC*, and *rbsK*, as determined by *in silico* SIFT analysis. With the exception of *ycgB* encoding a protein with unknown function, these genes all encode proteins predicted to be involved in the transport or metabolism of amino acids and carbohydrates (Table 3.1).

Table 3.1 The nonsynonymous mutations separating the MRCA of all Bhutanese *Shigella sonnei* from other CenAsiaIII clade strains

Gene	**Mutation**	**Function**	**SIFT score** *
cusS	G343S	sensory histidine kinase of CusSR, regulating the expression of CusCFBA operon to confer resistance to high copper and silver concentrations	0.77
ycgB	Y125H	unknown function	0
mppA	G344R	periplasmic binding component of murein tripeptide in oligopeptide transport	0
hisC	P98S	histidinol-phosphate aminotransferase - catalysing the biosynthesis of histidine	0.03
eutL	A22S	carboxysomal shell protein, structural component of proteinaceous microcompartment for ethanolamine ammonia lyase	0.5
hycC	G538D	component of HycBCDEFG, hydrogenase component (hydrogenase 3) of formate hydrogenlyase. HycC encodes an extremely hydrophobic protein with homology to NADH:ubiquinone oxidoreductase	0
sdaB	A407S	L-serine deaminase II, catalyzing the conversion of serine into pyruvate and ammonia	0.33
metC	A2T	Cystathionine B-lyase and cysteine desulfhydrase, catalyzing the conversion cystathionine to homocysteine.	0.7
rbsK	A50T	Ribokinase responsible for metabolism of D-ribose	0
Gene	**Mutation**	**Function**	**Known effect**
gyrA	D87G	DNA gyrase, subunit A	Resistance to fluoroquinolones
parC	S80I	DNA topoisomerase IV subunit A	Resistance to fluoroquinolones

* SIFT scores ≤0.05 indicate significant effect on protein function based on SIFT analysis

3.3.4 The accessory genome in Bhutanese *Shigella sonnei*

The composition of the accessory genome can be used to indicate the capacity of the host species to sample the local gene pool and also to highlight themes in the patterns of acquisition that may be under selection. I performed a comparative gene content analysis of Bhutanese and non-Bhutanese *S. sonnei* belonging to the CenAsiaIII clade. These data revealed events that appear to have been pivotal in evolution and spread of the CenAsiaIII clade. Notably, all the isolates of this clade possessed a specific class 2 integron In2 (*dfrA*1, *sat*1), conferring resistance to trimethoprim and streptothricin. These strains additionally gained the *S. sonnei* spA plasmid carrying *sul2*, *strAB* and *tetRA* genes, which further expanded the spectrum of antimicrobial resistance to incorporate sulphonamides, streptomycin and tetracycline resistance, respectively (Fig 3.1B). This resistance gene profile concurs with the observation made in the previous global *S. sonnei* study (Holt *et al.*, 2012). It is of note that the class 2 integron of the CenAsiaIII clade differs from that of Global III in the absence of *aadA1*, which confers resistance to aminoglycosides. Moreover, it was clear that isolates within this clade have, over time, accumulated chromosomal mutations conferring FQ resistance (Fig 3.1B). From the lineage-specific phylogeny, I predict that the first mutation in *gyrA* (S83L) occurred before 1996 (95% HPD; 1994 – 1998.5), and supplementary mutations in *parC* (S80I) and secondary mutations in *gyrA* (D87G), leading to high-level FQ resistance, were fixed in the Bhutanese clade prior to its expansion and diversification in 2006 (95% HPD; 2005 – 2008) (Fig 3.3).

A further pivotal event in the emergence of this CenAsiaIII clade prior to its introduction into Bhutan was the fixation of the previously described E-type colicin plasmid pSSE3 (with 100% homology in DNA sequence). This E-type colicin has been shown to be antagonistic against a variety of diarrhoeagenic *E. coli*, but not against commensal *E. coli* (Calcuttawala *et al.*, 2015). Plasmid pSSE3, which was found in all Bhutanese isolates (Figure 3.1B), also carries an immunity function that protects the host cell against the activity of the corresponding colicin. Sequence analysis of pSSE3 indicated that the colicin biosynthesis cluster was a novel recombinant hybrid with an undetermined spectrum of activity (Fig 3.4). On closer examination, the first and second domains (translocation and receptor-binding) of the colicin structural gene (*cea*) and the lysis gene (*cel*) exhibited substantial amino acid sequence identity (98%) to that of colicin E7 (ColE7). The third domain (cytotoxic) encoded by *cea* was most closely related to that of ColE6 (94% protein sequence identity). However, the product of the colicin immunity gene (*cei*) exhibited only 84% protein sequence identity to that of colicin E6.

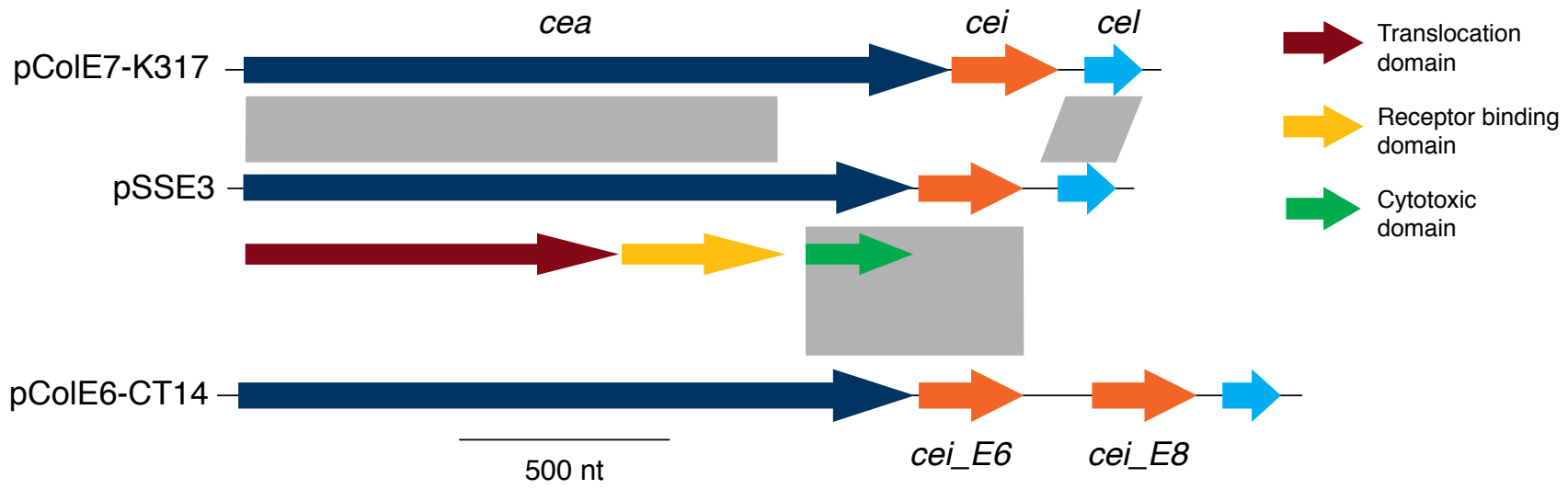


Figure 3.4 Novel recombinant hybrid colicin biosynthesis cluster encoded on pSSE3

Schematic representation of the colicin biosynthesis cluster on plasmid pSSE3 in comparison with colicin gene clusters of pColE7-K317 and pColE6-CT14. Grey vertical bars show homology in DNA sequences (with identity >80%). The colicin structural gene (*cea*), immunity gene (*cei*) and lysis gene (*cel*) are annotated as shown. The three domains of colicin structural gene are shown the key.

Genetic material that was acquired sporadically across the CenAsiaIII clade within the Bhutanese isolates included the ESBL gene, *bla*$_{\text{CTX-M-15}}$, which was independently acquired by two *S. sonnei* on different incompatibility plasmids (IncK and IncI1), conferring resistance to 3rd generation cephalosporins (Fig 3.3). The plasmid replication and transfer region of the IncK plasmid shared substantial sequence homology with that of the enteroaggregative *E. coli* (EAEC) IncI1 plasmid, pSERB1. This plasmid in the Bhutanese *S. sonnei* incorporated the type IV pilus biosynthesis cluster, *pilL-V* (Dudley *et al.*, 2006). For the *S. sonnei* containing the IncK plasmid, the *bla*$_{\text{CTX-M-15}}$ gene was encoded along with the FQ resistance enhancement gene, *qnrS1*, on an IS26 transposon. The *S. sonnei* IncI1 plasmid (denoted pHUSEC2011-1 like) was almost identical (≥99% DNA sequence identity) to the replication, conjugation and a type IV pilus biosynthesis cluster, *pilL-V*, in pHUSEC2011-1. The pHUSEC2011-1 like plasmid was also independently acquired by six additional strains, although in these cases the *bla*$_{\text{CTX-M-15}}$ gene was not co-transferred (Fig 3.3).

Lastly, nine *S. sonnei* strains isolated from cases occurring in June/July 2011 acquired an IncB-O plasmid, sharing homology with the MDR plasmid pHUSEC41-1. This was originally identified in an older O104:H4 enteroaggregative Shiga toxin-producing *E. coli* (HUSEC41), which included a type IV pilus biosynthesis cluster *pilL-V* identical to that found on plasmid pSERB1 (Künne *et al.*, 2012). I further identified eleven *S. sonnei* strains harbouring a plasmid with homology to *Salmonella enterica* plasmid pSH146_65, additionally carrying a distinct type IV pilus cluster *pilL-V* (Han *et al.*, 2012).

3.4 Discussion

Here I have used WGS to define a population of *S. sonnei* causing disease in the south Asian land-locked nation of Bhutan. These strains belong to an expansion of Lineage III, which was named the CenAsiaIII clade and introduced into Bhutan around 2006. In contrast to recent observations from Vietnam, the sequenced *S. sonnei* in Bhutan demonstrated no evidence of a single clonal expansion. This is perhaps due to the short sampling timeframe or other factors such as the lack of a definitive selective sweep event. However, the evolution history of CenAsiaIII clade *S. sonnei* resembles the two initial sweep events observed in Vietnam (associated with fixation of a colicin plasmid and mutation(s) in *gyrA*) (Holt et al. 2013), showing that the two geographically and phylogenetically unrelated *S. sonnei* populations followed a similar evolutionary trajectory. After its introduction to Bhutan, my data predict that these *S. sonnei* had access to a substantial gene pool, as indicated by the sporadic acquisition of ESBL plasmids and type IV pili. The gains of ESBL plasmids, though infrequent, could potentially become the next selective sweep for the Bhutan's *S. sonnei* population as observed in Vietnam, given the selection path toward resistance is maintained. The characteristically thin type IV pilus has been shown to improve the efficiency and specificity of conjugation in liquid environments (Kim and Komano, 1997), and also facilitates adherence, interbacterial contact and twitching motility (Berry and Pelicic, 2015). Although this structure is known to be widespread in prokaryotes, its impact on the lifestyle of *Shigella* is not well described and requires further experimental investigation (Holt *et al.*, 2013; Rohmer *et al.*, 2014).

It was possible to use WGS to identify the major events during the evolution of this lineage prior to, and after, its introduction into a new setting. My data imply that the acquisition of a class 2 integron and a spA plasmid, together with the rapid accumulation of mutations associated with FQ resistance, rendered the Bhutanese *S. sonnei* MDR by 2006. Further, an FQ resistant subclade within the CenAsiaIII clade was predicted to have arisen between 1999 and 2006. In addition, the fixation of a novel recombinant colicin E plasmid (pSSE3) may have offered this group of pathogens a competitive advantage over other enteric bacteria, similar to the colicin associated selective sweep observed in *S. sonnei* in Vietnam (Holt *et al.*, 2013). These data suggest that whilst antimicrobial resistance is clearly important, it is not sufficient in itself for driving the spread of new clones and competitive functions appear to be key themes in the spread of MDR *S. sonnei*.

Extensive literature additionally suggests that international travel may have resulted in the dissemination of CenAsiaIII clade's *S. sonnei* beyond Bhutan and the remainder of the Indian subcontinent. *S. sonnei* isolates with similar PFGE patterns to the strains in this analysis have been recovered from cases of travel associated shigellosis in Ireland (De Lappe *et al.*, 2015) and in a *Shigella* outbreak among MSM in Canada (Gaudreau *et al.*, 2011). The *S. sonnei* isolated from both instances were found to be ciprofloxacin resistant. Studies conducted in Japan have established a strong association between this pulsotype and travel to India, and increasing prevalence of nalidixic acid resistant *S. sonnei* has been observed since 2001 (Uchimura, Kishida and Koiwai, 2001; Izumiya *et al.*, 2009). Shigellosis surveys in Iran conducted from 2008 to 2012 have repeatedly

shown that this characteristic Asian pulsotype is prevalent and endemic, and that these strains also possess a similar class 2 integron (*dfrA1*, *sat1*) but exhibit variability in their ciprofloxacin susceptibility phenotype (Tajbakhsh *et al.*, 2012; Eftekhari *et al.*, 2013; Alizadeh-Hesar, Bakhshi and Najar-Peerayeh, 2015). The incidence of ciprofloxacin resistant *Shigella* in Asia and Africa has risen dramatically, from 3.9% in 2001-2003 to 29.1% in 2007-2009 (Gu *et al.*, 2012). The potential emergence and dissemination of FQ resistant *S. sonnei* poses a substantial threat to public health, including low shigellosis burden areas in developed nations, as again substantiated by the recent FQ resistant *S. sonnei* outbreak in the U.S.A. (Bowen *et al.*, 2015). Taken together, these data strongly suggest that the CenAsiaIII clade of *S. sonnei* is widely present across South Asia and beyond and may represent an emerging threat to public health.

4 The emergence and global spread of fluoroquinolone resistant *Shigella sonnei*

A part of the work described in Chapter 4 has been published during my DPhil, under an article in the journal PLoS Medicine named:

Chung The, H., Rabaa, M. A., Pham Thanh, D., Lappe, N. De, Cormican, M., Valcanis, M., Howden, B. P., Wangchuk, S., Bodhidatta, L., Mason, C. J., Nguyen, T. N. T., Thuy, D. V., Thompson, C. N., Lan, N. P. H., Vinh, P. V., Thanh, T. H., Turner, P., Sar, P., Thwaites, G., Thomson, N. R., Holt, K. E. and Baker, S. (2016) 'South Asia as a reservoir for the global spread of ciprofloxacin resistant *Shigella sonnei*', *PLoS Medicine*, 13(8), pp. 1–12. doi: 10.1371/journal.pmed.1002055.

4.1 Abstract

Antimicrobial resistance is a major issue in *Shigella*, particularly as a specific MDR lineage of *Shigella sonnei* (lineage III) is becoming globally dominant. Ciprofloxacin is a recommended treatment for shigellosis. However, ciprofloxacin resistant *S. sonnei* are being increasingly isolated in Asia and sporadically reported on other continents. I hypothesized that Asia is a primary hub for the recent international spread of ciprofloxacin resistant *S. sonnei*. Whole-genome sequencing was performed on a collection of more than 300 ciprofloxacin resistant *S. sonnei* isolated within the last decade from endemic Asian countries (Vietnam, n = 22; Bhutan, n = 71) and developed regions (Australia, n = 46; Ireland, n = 16; England, n = 82; France, n = 72). The recent

evolutionary history of these organisms was reconstructed in combination with their temporal information and geographical locations. I found that all sequenced ciprofloxacin resistant *S. sonnei* formed a single clade within a Central Asian expansion of lineage III. Furthermore, the data here show that resistance to ciprofloxacin within *S. sonnei* is likely globally attributed to a single clonal emergence event, encompassing sequential *gyrA*-S83L, *parC*-S80I, and *gyrA*-D87G mutations. Geographical data predict that South Asia is the likely primary source of these organisms, which are being exported to Southeast Asia and Europe. Using Bayesian phylogenetic inference, I estimated that this ciprofloxacin resistant *S. sonnei* clone diverged from a common ancestor in early 2007. This study suggests that a single clone, which is widespread in South Asia, is likely driving the current intercontinental surge of ciprofloxacin resistant *S. sonnei* and is capable of establishing endemic transmission in new locations.

4.2 Introduction

Although shigellosis is typically self-limiting, antimicrobial treatment is used to prevent complications, reduce diarrhoeal discharge and restrict faecal shedding (Bennish, 1991; Vinh *et al.*, 2000). Consequently, resistance to antimicrobials limits treatment options, placing vulnerable individuals suffering from shigellosis at increased risk of complications and increasing the likelihood of protracted faecal shedding. One of the current recommended first-line treatments for shigellosis is the FQ, ciprofloxacin (WHO, 2005). The FQs target the DNA gyrase, a type II topoisomerase that is essential for bacterial DNA replication and transcription (Ruiz, 2003).

Present-day international epidemiology of the various *Shigella* species is particularly intriguing, as *S. sonnei* is replacing *S. flexneri* as the most common cause of shigellosis worldwide; this pattern is accentuated in regions undergoing rapid economic development, where *S. flexneri* dominated as recently as a decade ago (Vinh *et al.*, 2009; Thompson, Duy and Baker, 2015). Antimicrobial resistance is an emerging global issue in *S. sonnei,* with MDR lineage III now dominating internationally. Given their common usage and broad spectrum of activity, resistance against the FQs is the most concerning. Since the first isolation of *S. sonnei* with reduced susceptibility to ciprofloxacin in Japan in 1993 (Horiuchi *et al.*, 1993), ciprofloxacin resistant *S. sonnei* have been increasingly reported throughout Asia (Nandy *et al.*, 2011; Ud-Din *et al.*, 2013; Ruekit *et al.*, 2014). Furthermore, public health laboratories in several non-Asian countries with low incidences of shigellosis have also reported the isolation of ciprofloxacin resistant *S. sonnei*, often from individuals reporting recent travel to locations with a high risk of shigellosis (Bowen *et al.*, 2015; De Lappe *et al.*, 2015; Kim *et al.*, 2015).

WGS has proven to be the gold standard for tracking the international dissemination of clonal bacterial pathogens (He *et al.*, 2013; Wong *et al.*, 2015), and the phylogenetic structure and spread of *S. sonnei* at both international and national levels have been deciphered (Holt *et al.*, 2012, 2013). Hypothesising that Asia was a hub for the recent global spread of ciprofloxacin resistant *S. sonnei,* I performed WGS and phylogenetic characterisation of more than 300 ciprofloxacin resistant *S. sonnei* isolated from within and outside Asia, aiming to explore the origins of this growing international epidemic. In addition, I aimed to elucidate the mutational differences in ciprofloxacin resistant and

sensitive *S. sonnei* under the presence and absence of FQ exposure *in vitro*, which allows us to comprehend the short-term evolution of the pathogen.

4.3 Results

4.3.1 Fluoroquinolone resistant *Shigella sonnei* in global context

I constructed a whole genome phylogeny of *S. sonnei*, incorporating sequences from 133 representative isolates from the previous global *S. sonnei* phylogeny study (Holt *et al.*, 2012) and a selection of 60 novel isolates from Vietnam, Cambodia, Thailand, Bhutan, Australia, and Ireland. These novel sequences were ciprofloxacin resistant (MIC >1 μg/ml or zone of inhibition < 19 mm) organism. The overall tree topology reflected the previously described global phylogenetic structure (Holt *et al.*, 2012), confirming the presence of four distinct lineages (I, II, III and IV); lineage III was the most commonly represented and the most widely geographically distributed (Fig 4.1). All ciprofloxacin resistant *S. sonnei* formed a single well-supported monophyletic clade within the Central Asian expansion of Lineage III (CenAsiaIII), an MDR group that is closely related but distinct from the Global III clade (Fig 4.1).

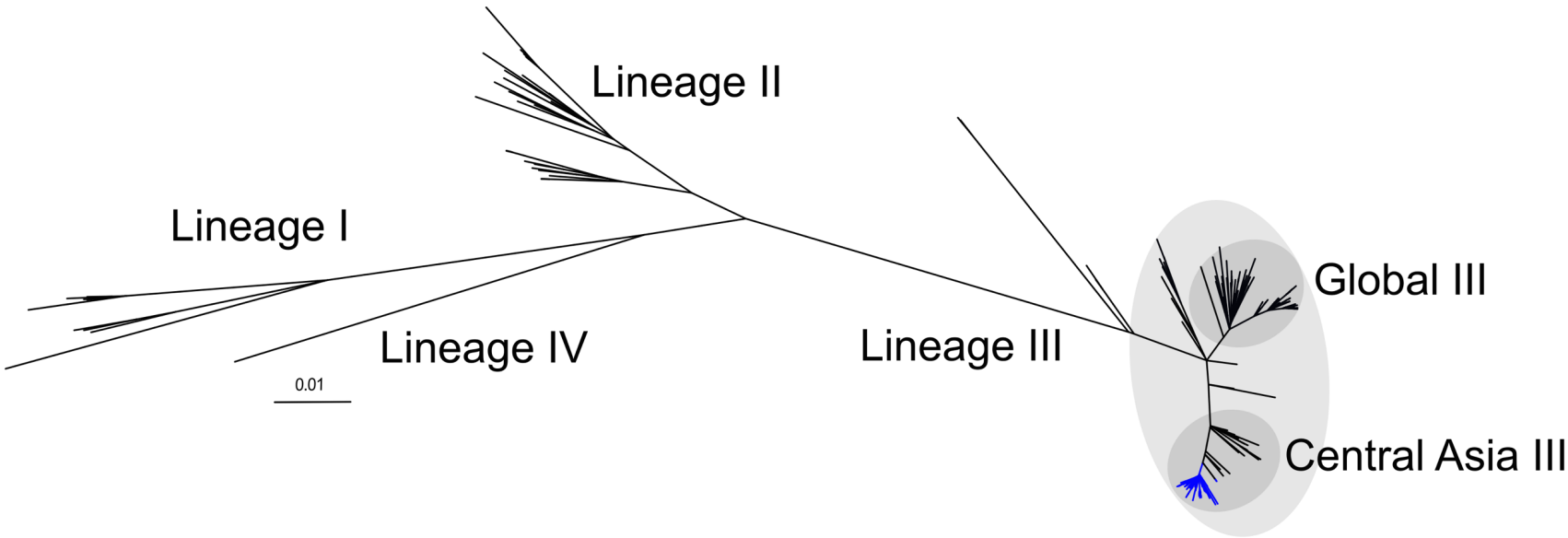


Figure 4.1 The phylogenetic structure of ciprofloxacin resistant *Shigella sonnei* in an international context. The figure shows an unrooted maximum likelihood phylogeny of 211 globally representative *S. sonnei*, including 60 sequences from ciprofloxacin resistant isolates (highlighted by the blue branches). Major lineages are indicated by numbers (I, II, III and IV) as defined in Holt et al. 2012, with clades Global III and Central Asia III within lineage III highlighted. Horizontal scale bar indicates the number of nucleotide substitutions per site.

4.3.2 The emergence of the fluoroquinolone resistant *Shigella sonnei* clone

For a more detailed phylogenetic investigation, I gathered a contemporary collection of 411 *S. sonnei* isolated internationally to enhance the analysis's power and representativeness (Table 4.1). Within this collection, 313 isolates exhibited phenotypic resistance to ciprofloxacin and were isolated since 2008 (with two isolates from Vietnam later excluded due to poor sequencing quality). The dataset includes *S. sonnei* isolated from endemic regions (Bhutan, Vietnam, Thailand, Cambodia) and from public health laboratories of developed regions (Ireland, Australia, England, France). For isolates sourced from non-endemic areas, the patients' recent travel histories were recorded to identify the original geographical sources of infections. Preliminary phylogenetic reconstruction showed that eleven isolates (among the 411), sourced from South and Southeast Asia (five of which showed reduced susceptibility to ciprofloxacin, MIC ranges 0.094 to 0.125), clustered outside the CenAsiaIII clade and were removed from further phylogenetic analyses. All the 311 examined ciprofloxacin resistant organisms belong to the previously characterized CenAsiaIII clade, which agrees with the previous analysis on a smaller dataset.

Table 4.1 Summary of sequences used to investigate the emergence of fluoroquinolone resistant *Shigella sonnei*. CIP: ciprofloxacin

Country	Total	In Cen-AsiaIII clade	CIP resistant isolates	Study or Institute Origin	Patient group	Region of recent travel history (N)	sequencing platform	CIP Susceptibility test
Bhutan	71	71	71	Diarrhoeal disease surveillance in JDWNRH, Thimphu, Bhutan (AFRIMS)	hospitalized children < 5 y old	NA	Illumina Hiseq 2000	Disk diffusion/E-test
Vietnam	24	22	22	Diarrhoeal disease surveillance in HCMC, Vietnam	hospitalized children < 5 y old	NA	Illumina Miseq	Disk diffusion
Thailand	8	8	1	Cross-sectional study of *S. sonnei* diversity in Southeast Asia	hospitalized children < 5 y old	NA	Illumina Hiseq 2000	Disk diffusion
Cambodia	1	1	1	Cross-sectional study of *S. sonnei* diversity in Southeast Asia	hospitalized children < 5 y old	NA	Illumina Hiseq 2000	Disk diffusion
Ireland	20	20	16	National *Salmonella*, *Shigella*, and *L. monocytogenes* Reference Laboratory, Galway, Ireland	Primarily patients with recent travel history	South Asia (9), Africa (1), Europe (1), No travel (5), Unknown (4)	Illumina Hiseq 2000	Broth microdilution

Australia	85	77	46	Microbiological Diagnostic Unit Public Health Laboratory in Melbourne, Australia	Primarily patients with recent travel history	Africa (2); America (1); Southest Asia (12); Middle East (2); Oceania (1); South Asia (60); Unknown (4)	Illumina Nextseq	Agar dilution
France	97	95	72	French National Reference Centre for *E. coli*, *Shigella*, and *Salmonella*, Paris, France	Primarily patients with recent travel history	Africa (4); America (1); Southeast Asia (4); South Asia (88)	Illumina Hiseq 2000	Disk diffusion
England	91	90	82	Gastrointestinal Bacteria Reference Unit, Public Health England, London, United Kingdom	Primarily patients with recent travel history	Africa (3); East Asia (1); Europe (13); Middle East (1); South Asia (73)	Illumina Hiseq 2000	Disk diffusion
Global references	14	11	0	NA	NA	NA	Illumina Hiseq2000	NA
	411	395	311					

In-depth phylogenetic reconstruction of 395 CenAsiaIII *S. sonnei* verified that almost all the FQ resistant *S. sonnei* (306/311) belonged to a single clone, with resistance conferred by mutations in the quinolone resistance-determining region (QRDR), with the sequential accumulations in *gyrA* (S83L), *parC* (S80I), and *gyrA* (D87G) (Fig 4.2). A single outlier isolate shares the *gyrA*-S83L and *parC*-S80I mutations, but harbours *gyrA*-D87N rather than a *gyrA*-D87G, and is within a closely related outgroup of the major resistant clone. In addition, mutations affecting *gyrA*-D87 appears to arise in at least four independent occasions, leading to reduced susceptibility to FQ. The clade could be further subdivided into two populations using BAPS, named Pop1 (FQ susceptible) and Pop2 (mostly FQ non-susceptible) (Fig 4.3). Within Pop2, the majority of organisms originated from South Asia (234/307). The CenAsiaIII population showed presence of minimal recombination (Fig 4.4). The mean nucleotide substitution rate of the CenAsiaIII clade was estimated to be 7.34×10^{-7} substitutions per site per year [95% HPD: 6.77×10^{-7} to 7.91×10^{-7}], using methods described in section 2.3.4. This suggests the clade diverged in the late 1980s [95% HPD: 1986-1990] (Fig 4.5). This timeline is in accordance with previous estimates from the global evolution of *S. sonnei* (Holt *et al.*, 2012). Two lineage-defining QRDR mutations (*gyrA* S83L and D87Y) arose independently within the mid-1990s (1995-1996), forming two major subclades with quinolone (i.e. nalidixic acid) resistance. The second QRDR mutation (*parC* S80I) arose in the *gyrA* S83L background between 2001-2005, and the third (*gyrA* D87G) defining the FQ resistant *S. sonnei* population occurred by 2007 (95% HPD: 2006.55 – 2007.62). Since its divergence, it appears that the FQ resistant *S. sonnei* population expanded rapidly, peaking in 2008, and subsequently maintained a relatively consistent size (Fig 4.6). The slight collapse in mid 2011 is likely a result of sampling bias since demographic inference is sensitive to sampling frequency.

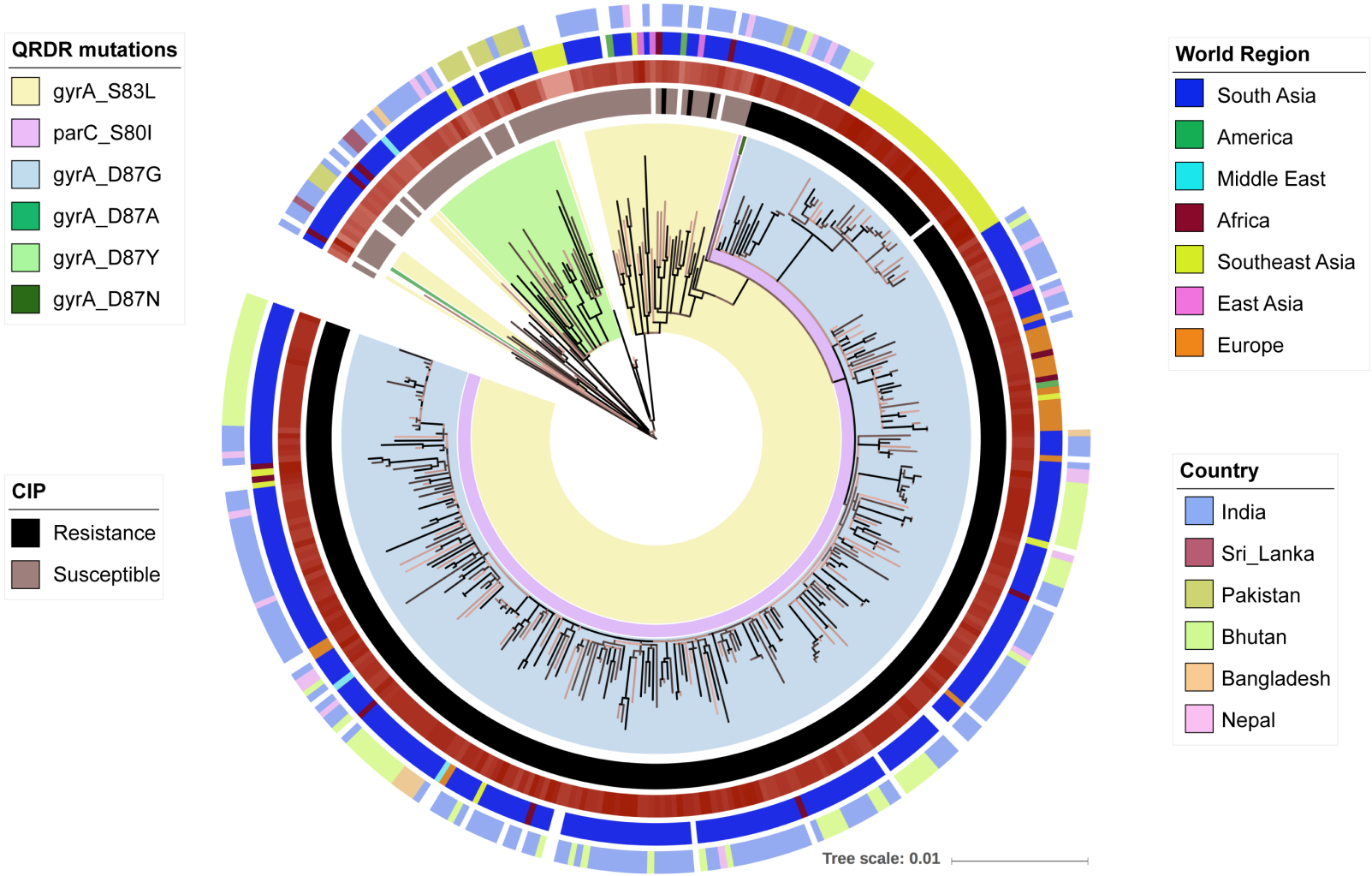

QRDR mutations
gyrA_S83L
parC_S80I
gyrA_D87G
gyrA_D87A
gyrA_D87Y
gyrA_D87N
CIP
Resistance
Susceptible
World Region
South Asia
America
Middle East
Africa
Southeast Asia
East Asia
Europe
Country
India
Sri_Lanka
Pakistan
Bhutan
Bangladesh
Nepal
Tree scale: 0.01

Figure 4.2 The phylogenetic structure of CenAsiaIII *Shigella sonnei*. The figure displays the maximum likelihood phylogeny of 395 *S. sonnei* sequences belonging to the CenAsiaIII clade. The tree is rooted on the most closely related sequence from the *S. sonnei* global sequencing study (Holt *et al.*, 2012). The branch colour scheme indicates the span of bootstrap support for the corresponding branch, ranging from low to high (light pink to black). The shading covered on the phylogeny indicates taxa harbouring quinolone resistance determining region (QRDR) mutations (see legend). The rings show the associated information for each taxon, from the innermost to the outermost in the following order: (1) Susceptibility testing for ciprofloxacin (black: resistance; brown: susceptible); (2) the year of isolation, ranging from 1996 to 2015 (light to dark red); (3) original geographical region of isolation; (4) where strain originates from South Asia, the specific country of isolation (see legends). The horizontal scale bar indicates the number of nucleotide substitutions per site.

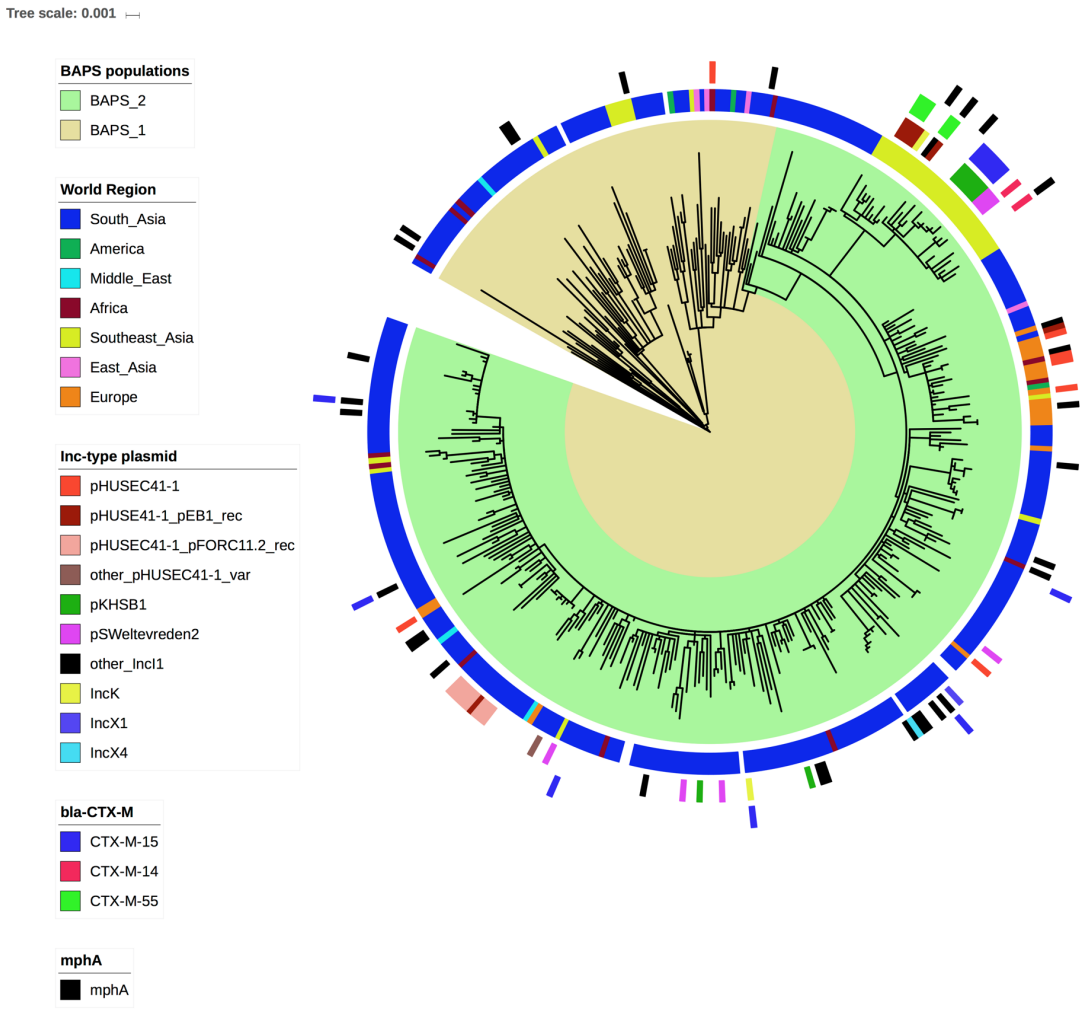


Figure 4.3 Plasmid profile and antimicrobial resistance genes among 395 CenAsiaIII *Shigella sonnei*. Figure shows the maximum likelihood phylogeny identical to that described in Figure 4.2. The shading over the phylogeny indicates the population clusters based on Bayesian Analysis of Population Structure (BAPS) hierarchical clustering (Pop1 and Pop2). The rings present the associated data for each taxon, detailed from the innermost to the outermost as following: (1) original geographical region of isolation (see legends); (2) Presence of plasmids; (3) Presence of Extended Spectrum Beta-Lactamase (ESBL) *bla*$_{\text{CTX-M}}$ genes; (4) Presence of *mphA*. The horizontal scale bar denotes the number of substitutions per site.

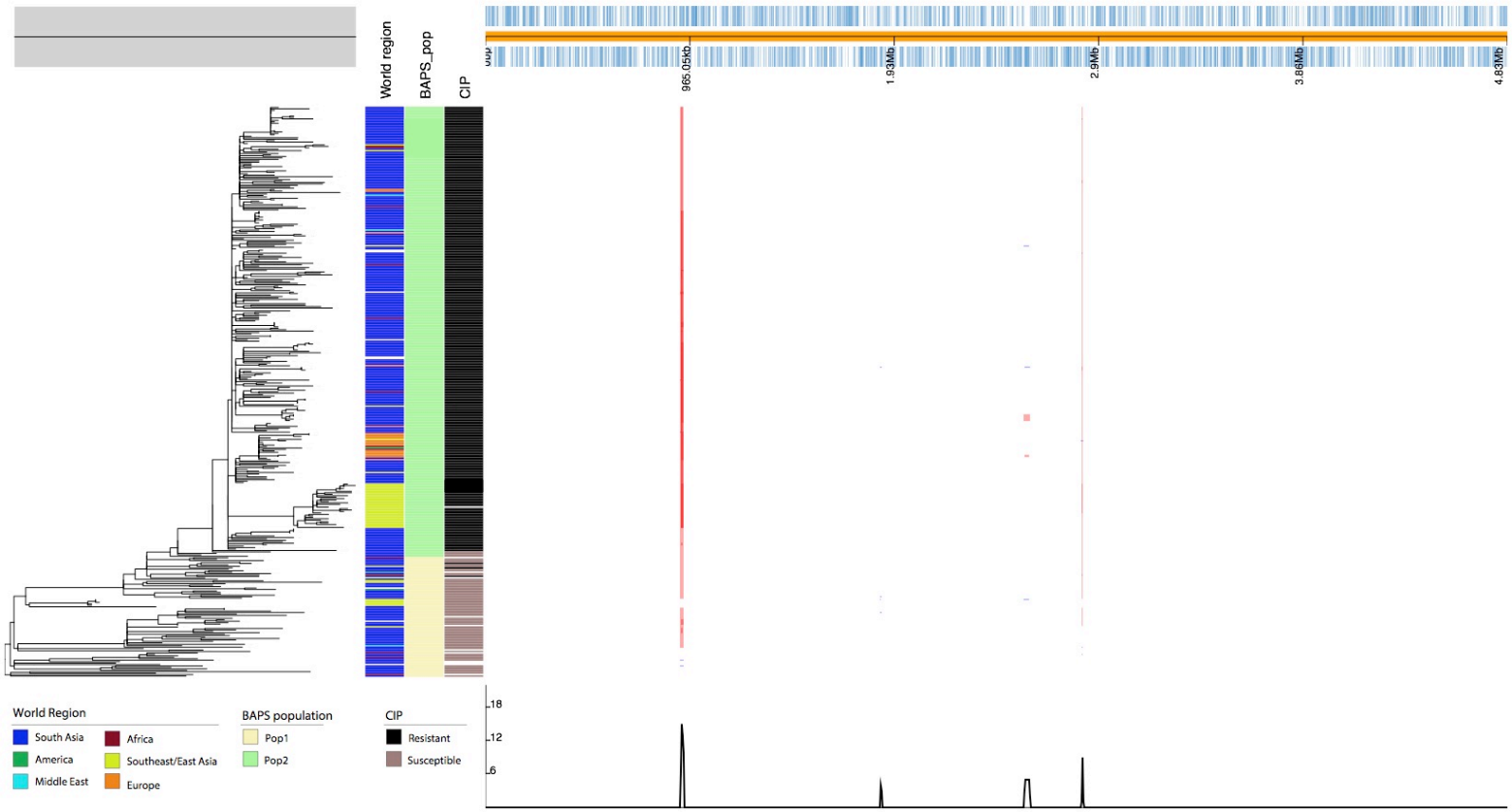


Figure 4.4 Recombination features within the genomes of 395 CenAsiaIII *S. sonnei* isolates

The phylogeny on the left is identical to the one described in Figure 4.2, with information regarding isolation source, population structure and ciprofloxacin susceptibility attached to each taxon (see Legends). The top right bar displays the genomic coordinates of *S. sonnei* reference Ss046, with coding sequences coloured in blue. Presumed recombination regions are detected within the genomes using Gubbins, which are highlighted in blue (occurring in single taxon) or red (shared between many monophyletic taxa).

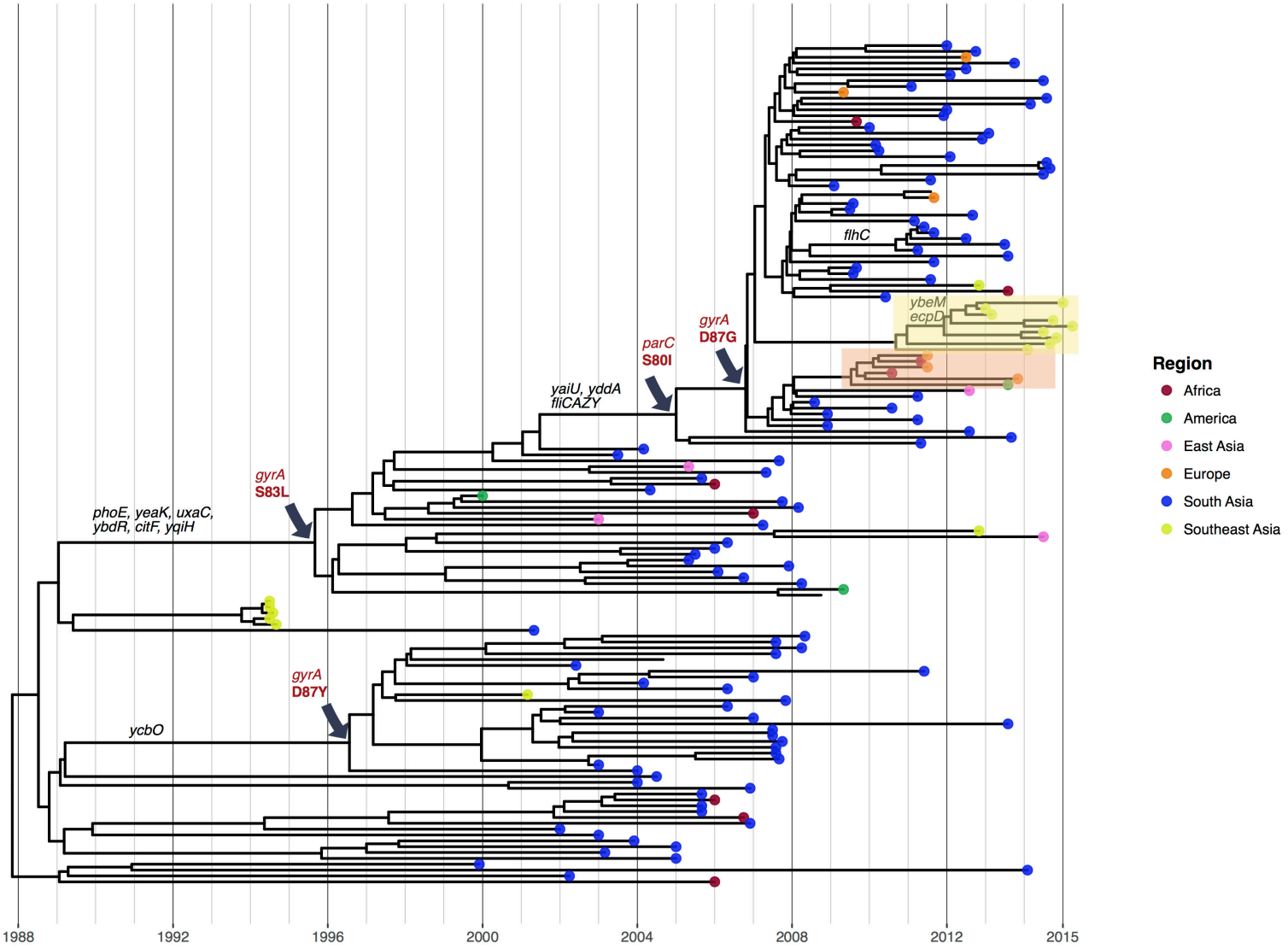

gyrA
S83L
parC
S80I
gyrA
D87G
gyrA
D87Y
yaiU, yddA
fliCAZY
phoE, yeaK, uxaC,
ybdR, citF, yqiH
ycbO
flhC
ybeM
ecpD
1988
1992
1996
2000
2004
2008
2012
2015
Region
Africa
America
East Asia
Europe
South Asia
Southeast Asia

Figure 4.5 Temporal phylogenetic reconstruction of CenAsiaIII *Shigella sonnei* overlaid with pseudogenes. Figure shows a maximum clade credibility phylogenetic reconstruction of 144 CenAsiaIII *S. sonnei*. Grey arrows designate branches on which certain QRDR mutations occurred. Lineage-defining pseudogenes and gene loss events are denoted above corresponding branches. Tip colours show the original geographical isolation region of taxa (see legends). The orange and yellow boxes highlight the European and Southeast Asian clonal expansion of CenAsiaIII respectively.

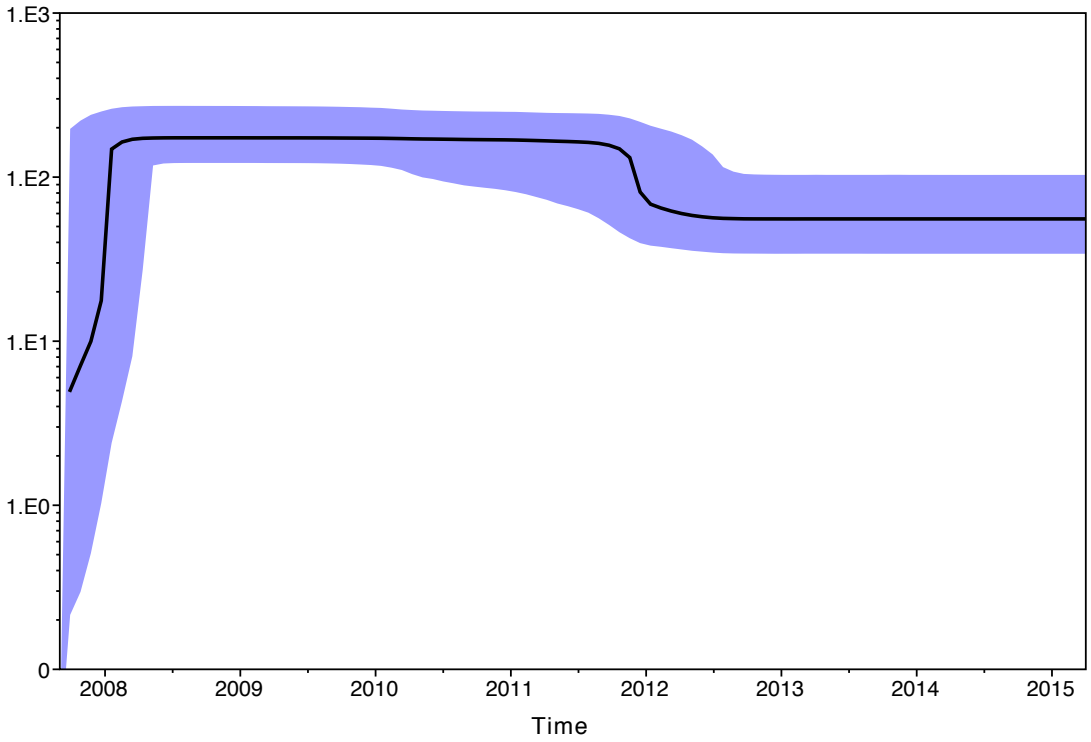


Figure 4.6 Bayesian skyline plot showing the change in effective population size of CenAsiaIII *Shigella sonnei*. The analysis was performed on 134 isolates of FQ resistant *S. sonnei*, using the TVM substitution model coupled with a strict molecular clock and Bayesian skyline demographic model.

4.3.3 Phylogeography and the origin of fluoroquinolone resistant *Shigella sonnei*

Two distinct regional diversification events from South Asia are evident, with one being introduced into Southeast Asia (Vietnam, Thailand, Cambodia) in 2011, and another further instigating sustained transmission within Europe from 2009 (Fig 4.5). Both are supported by branches with 100% bootstrap values. Notably, despite dense sampling, there was no immediate out-group for the Southeast Asian subclone. Detailed examination showed that there were 18 genomically-dispersed mutations specific to this Southeast Asia subclone, confirming its divergence from the trunk of phylogeny. The isolates from this subclone also constitute ones isolated from Australian and European patients, emphasizing that international travel potentially leads to secondary introductions to shigellosis non-endemic regions. The second subclone, comprised mainly of organism from Europe (Ireland, Italy, Germany, and Spain), shares its most closely related out-groups with South Asian sequences, demonstrating its likely emergence from this region. I next mapped the isolates' geographical sources to the phylogeny and showed that South Asia (including India, Bhutan, Nepal, etc.) was the most predominant representation (Fig 4.2). To avoid bias arising from unequal samplings, I employed stochastic mapping to evaluate the most likely evolutionary origin of CenAsiaIII clade, with equal sampling from each geography. It was found that South Asia was the key hub for all intercontinental transmissions, with CenAsiaIII *S. sonnei* being exported from South Asia to Africa, Southeast/East Asia, and Europe on an average of 7.65 (IQR: 5 – 9), 3.47 (IQR: 2 – 5), and 3.8 (IQR: 3 – 4) independent events, respectively (Fig 4.7A). Approximately half of the evolutionary time of CenAsiaIII has been in South Asia,

further confirming that the clade has circulated and evolved mainly in this region (Fig 4.7B).

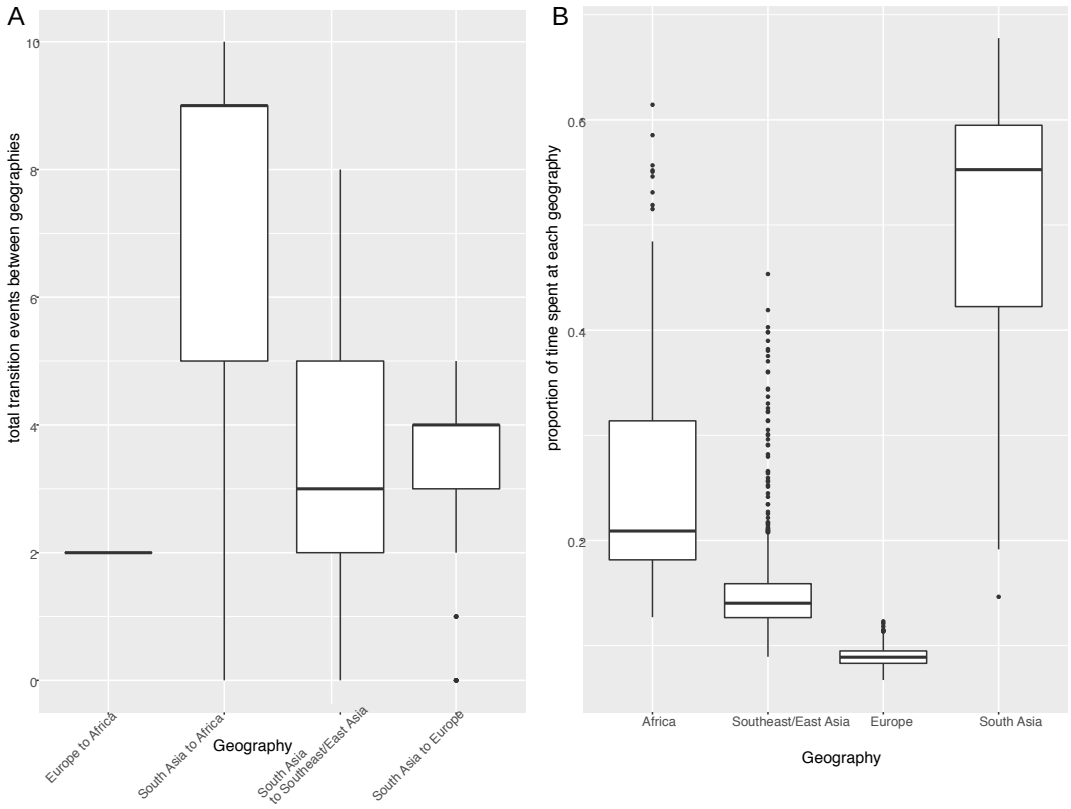


Figure 4.7 Geographical spreads of CenAsiaIII *Shigella sonnei.* Stochastic mapping was used to estimate (A) The number of transitions between different geographical states: Africa, Southeast/East Asia, Europe, and South Asia. Transitions between states that are less than one are considered insignificant and not represented in the plot. The boxplots summarize results of 966 successful sub-samplings of the phylogeny described in Figure 4.2, and stochastic mapping was performed for each with 100 simulations. (B) The proportion of time spent in each geographical state using the same sub-sampling and stochastic mapping analyses detailed in (A).

4.3.4 Accessory genome investigation

Aside from FQ resistance, it is of additional relevance to characterize the AMR within the CenAsiaIII *S. sonnei* population. Genes encoding resistance to older antimicrobials, such as *strAB* (streptomycin), *tetA* (tetracycline), *dfrA1-sul2* (co-trimoxazole) were commonplace within the clade, as previously reported (Holt *et al.*, 2012; Chung The *et al.*, 2015). Other notable AMR markers include genes belonging to the ESBL $bla_{\text{CTX-M}}$ family; these were particularly enriched in the Southeast Asian expansion (Fig 4.3). Despite the over-representation of South Asian isolates, $bla_{\text{CTX-M}}$ gene was found only sporadically in organisms arising from this region. When I account for isolates sourced from regional surveillance studies, including Bhutan and Vietnam, there was a stark contrast in the distribution of ESBL genes. Three $bla_{\text{CTX-M}}$ variants have been incorporated in different plasmid structures and acquired on separate occasions in Vietnam (Fig 4.3). While $bla_{\text{CTX-M-55}}$ co-exists with an IncB-O plasmid (recombinant of pEB1 and pHUSEC41-1) in five Vietnamese isolates, the $bla_{\text{CTX-M-15}}$ variant has been co-transferred with the IncI1 plasmid pKHSB1 in another six Vietnamese isolates. This latter plasmid became ubiquitous within the indigenous Global III-clade *S. sonnei* population from 2006, highlighting the promiscuousness of a specific plasmid among different *S. sonnei* lineages (Holt *et al.*, 2013). I additionally observed four independent instances of *mphA* acquisition within the Vietnamese expansion, two of which are co-transferred with $bla_{\text{CTX-M}}$ genes (Fig 4.3). Significantly, this combination of mutations and AMR genes renders these organisms resistant to all first line antimicrobials (FQ, 3rd generation cephalosporins, and macrolides).

I catalogued the presence of plasmids in all CenAsiaIII *S. sonnei* and found that IncB-O and IncI1 plasmids were the most common (Fig 4.3). A detailed plasmid comparison found that all IncB-O plasmids were closely related to pHUSEC41-1 or its recombinant variants. These plasmids exhibited an intimate association with geographical location, with pHUSEC41-1 mostly circulating in Europe, while the pHUSEC41-1-pEB1 and pHUSEC41-1-pFORC11.2 recombinants circulating in Vietnam and Bhutan, respectively (Fig 4.3). Alternatively, the IncI1 plasmids displayed more diversity with respect to their origins, with plasmid structures similar to those described in *Salmonella* spp. (14/44), *E. coli* (13/44), *S. sonnei* (11/44), and *K. pneumoniae* (3/44). Most notably, two Vietnamese sub-populations separately harboured pKHSB1 (*S. sonnei*) and pSWeltevreden2, originally described in *Salmonella* Weltevreden (Fig 4.3). These data demonstrated that once this variant of *S. sonnei* is introduced into a location, it has access to a wide gene pool shared among the indigenous Enterobacteriaceae. In addition, I previously reported that a novel recombinant colicin plasmid (pSSE3) entered into the CenAsiaIII clade (Chung The *et al.*, 2015). Here, I found that this plasmid has been maintained in this population since 1995 (Fig 4.8). This plasmid likely grants CenAsiaIII *S. sonnei* a competitive advantage against other *Shigella* and *E. coli* through colicin-mediated killing activity (Calcuttawala *et al.*, 2015). Furthermore, various clusters of different colicin families have been sporadically added across the phylogeny, with *colIa* and *colIb* integrated into the IncI1 plasmid backbones on thirteen occasions. These data showed that maintenance and enrichment of colicin may facilitate *S. sonnei* in forming a successful clonal expansion.

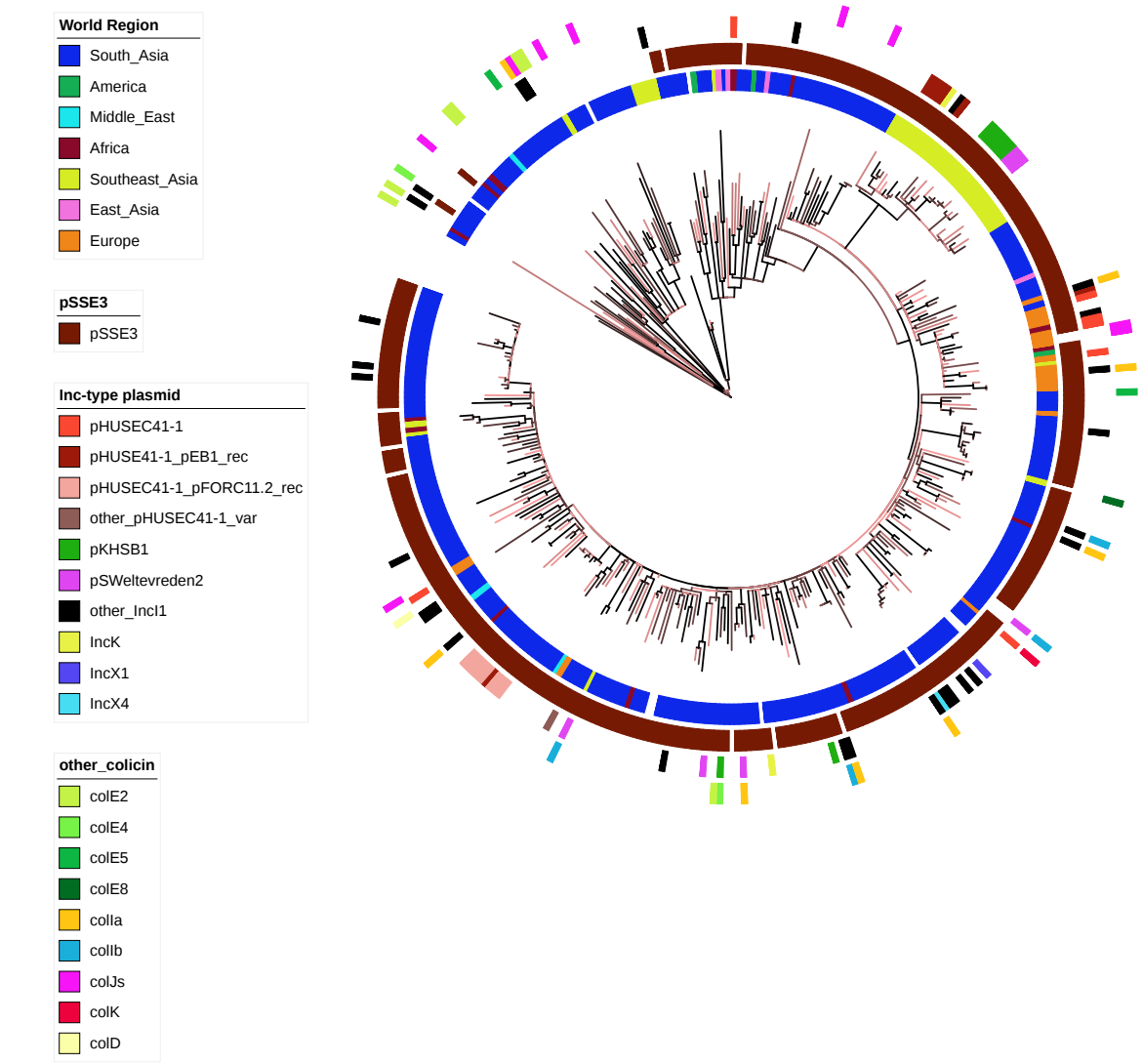


Figure 4.8 Plasmid profile and colicin families present in 395 CenAsiaIII *Shigella sonnei*. Figure shows the maximum likelihood phylogeny identical to that described in Figure 4.2. The rings present the associated data for each taxon, detailed from the innermost to the outermost as following: (1) original geographical region of isolation (see legends); (2) Presence of colicin plasmid pSSE3; (3) Presence of other plasmids; (4) Presence of colicin biosynthesis clusters other than that found on pSSE3. The horizontal scale bar denotes the number of substitutions per site.

4.3.5 Reductive evolution in CenAsiaIII *Shigella sonnei*

Reductive evolution plays an important role in shaping the long term evolution of *Shigella* (Prosseda *et al.*, 2012), but this phenomenon has not been sufficiently examined in a newly evolved population. Here I attempted to detail gene loss and pseudogene formation in the CenAsiaIII clade. In comparison to the Ss046 reference genome, *S. sonnei* accumulated 14 pseudogenes before the formation of CenAsiaIII (Table 4.2). Since 1988 (the estimated emergence of CenAsiaIII) a further thirteen lineage-defining mutations have arisen (Figure 4.5). The majority of these mutations are associated with long branches, potentially leading to multiple bottlenecks. These mutations are associated with genes that encode transporters (*araH*; arabinose, *ycbO*; aliphatic sulfonate, and other substrate nonspecific transporters *yjcR*, *phoE*, and *yddA*), and oxidation-reduction processes (*ydhV*, *ybeQ*, *mdaA*, and *ybdR*). Specifically, an inactivation in *mdaA* (*nfsA)* is known to result in nitrofurans resistance (Whiteway *et al.*, 1998). Further mutations potentially disrupt certain metabolic abilities, such as in the metabolism of glucuronate (*uxaC*) and citrate (*citF*). A truncation in *yeaK*, encoding for a deacylase capable of correcting tRNAs mischarged with serine/threonine (ProXp-ST1), has been shown to negatively impact on the translation fidelity (Liu *et al.*, 2015). The most distinctive mutation was the obliteration of the flagellin monomer and its transcription unit, *fliCAZY*. Through genomic comparison, I speculate this occurred in a stepwise fashion, first with an IS630 insertion into *fliC* at a specific target sequence (5'-CTAG-3') prior to 1988 (Tenzen and Ohtsubo, 1991). This IS element, in combination with an additional IS630 downstream of *fliY*, may act as an excision hotspot, leading to the deletion of the entire operon. Analysis of the pan-genome demonstrated that the spontaneous deletion of this

region has occurred on multiple independent occasions, including one leading to the formation of FQ resistant population (Pop2) prior to 2005. The *fliC* gene encodes for flagellin, a highly immunogenic protein and the monomer for flagellar assembly while *fliA* encodes for the flagellum-specific sigma factor ($E\sigma^{28}$). This genomic deletion leads to irreversible removal of the whole flagellar assembly, potentially explaining the subsequent inactivation of the redundant flagellar synthesis master regulator *flhC* in a Bhutanese *S. sonnei* subclone by 2011 (Fig 4.5). Cytokine profiling found that IL-1Ra is significantly lower in plasma of patients infected with FQ resistant *S. sonnei* (p=0.0338, logistic regression) compared to those infected with the susceptible counterpart (Fig 4.9). Monomeric flagellin is known to induce secretion of the anti-inflammatory IL-1Ra in intestinal epithelia and macrophages via its recognition by Toll-like receptor-5 (TLR5) (Carvalho *et al.*, 2011). These suggest that flagellin loss in FQ resistant *S. sonnei* may directly alter the infected individual's immune response.

Table 4.2 List of lineage-defining pseudogenes and gene loss in 395 CenAsiaIII *Shigella sonnei*.

Gene	Mode of inactivation	GO-function	No. of isolates	Estimated time of activation
ydhV	Disruption by IS600	Putative oxidation-reduction process	395	Since MRCA of CenAsiaIII
asmA	stop codon	Regulation of protein targeting to membrane	395	Since MRCA of CenAsiaIII
hofC	stop codon	Pilus organization	395	Since MRCA of CenAsiaIII
ybhM	stop codon	Inner membrane	395	Since MRCA of CenAsiaIII
yohL	stop codon	transcription regulation of cobalt and nickel homeostasis	395	Since MRCA of CenAsiaIII
ybeB		regulation of translation	395	Since MRCA of CenAsiaIII
araH	Disruption by IS600	arabinose transport	395	Since MRCA of CenAsiaIII
ydeK	Disruption by IS4	Membrane autotransporter	395	Since MRCA of CenAsiaIII
pepE	frameshift	Proteolysis	395	Since MRCA of CenAsiaIII
ybeQ	frameshift	oxidative stress response	395	Since MRCA of CenAsiaIII
ybeV	Inframe insertion (+1aa)	Regulation of ATPase activity	395	Since MRCA of CenAsiaIII
yjcR	Disruption by IS1S	Sulfur drug transmembrane transport	395	Since MRCA of CenAsiaIII
fliC	Disruption by IS630	Flagellin (cell motility)	395	Since MRCA of CenAsiaIII
mdaA	Disruption by IS1	oxidation-reduction	395	Since MRCA of CenAsiaIII
phoE	Disruption by IS2	Transport	341	Since 1996 with *gyrA* S83L
yeaK	Disruption by IS	Positive regulation of translational fidelity	341	Since 1996 with *gyrA* S83L

uxaC	stop codon	Glucuronate metabolism	341	Since 1996 with *gyrA* S83L
ybdR	stop codon	Oxidation-reduction	341	Since 1996 with *gyrA* S83L
citF	frameshift	Citrate metabolism	341	Since 1996 with *gyrA* S83L
yqiH	frameshift	Pilus organization	341	Since 1996 with *gyrA* S83L
fliCAZY	Deletion	Flagellum (cell motility)	334	Before *parC* S80I in 2005 + independent sporadic losses
yddA	Disruption by IS	Transport	309	Before *parC* S80I in 2005
yaiU	inframe deletion (-2aa)	Transport	309	Before *parC* S80I in 2005
flhC	stop codon	regulation of flagellum assembly	20	Bhutanese subclade
ybeM	Stop codon	Nitrogen compound metabolism	20	Southeast Asian expansion
ecpD	Stop codon	Pilus organization	20	Southeast Asian expansion
ycbO	Stop codon	Transport	31	BAPS-pop1, since *gyrA* D87Y

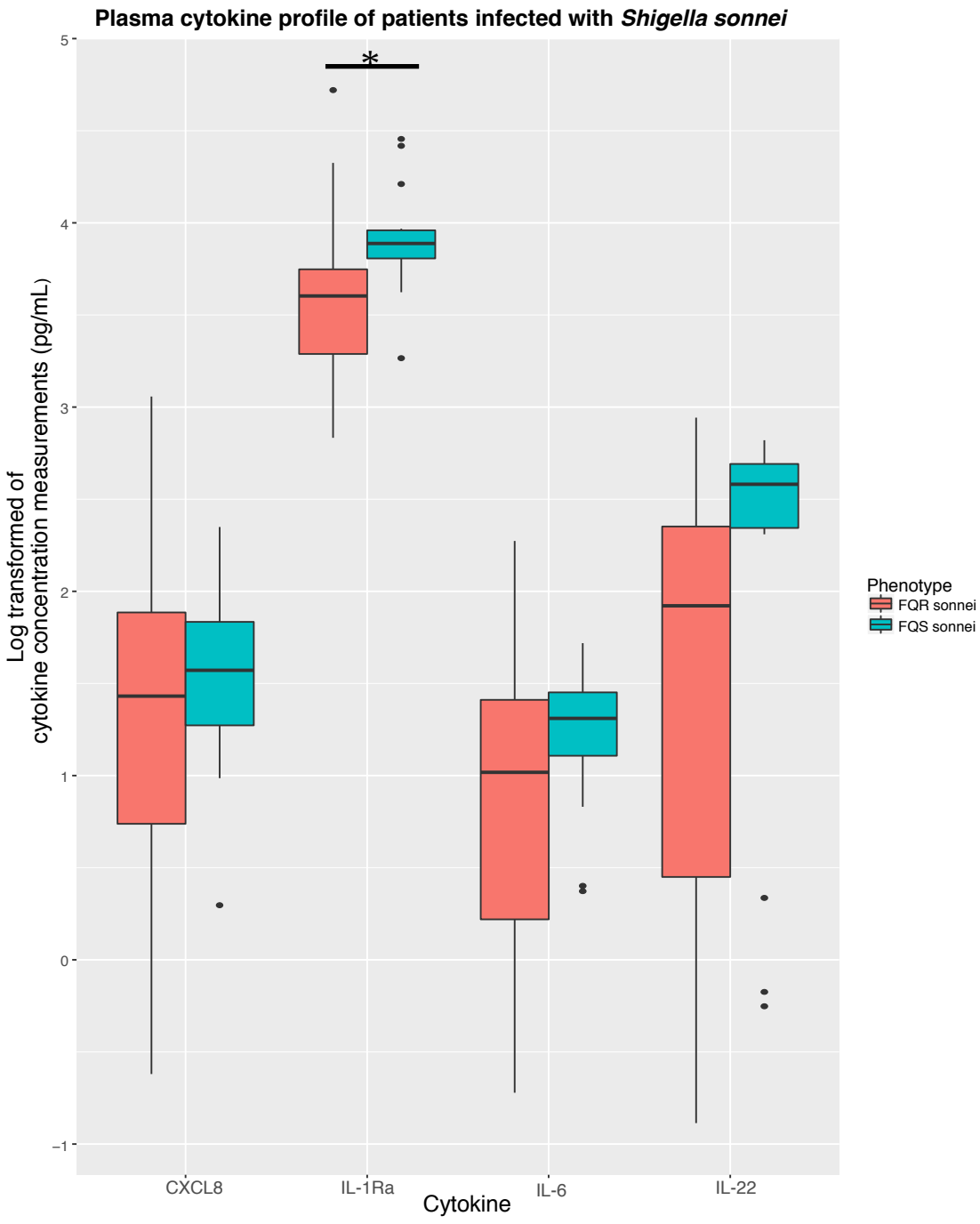


Figure 4.9 Cytokine measurements from plasma of patients infected with *Shigella sonnei*.

I next screened for pseudogenes formed by substitutions or small indels among the 393 CenAsiaIII *S. sonnei* to detect signals of convergent evolution. This analysis revealed that *rpoS* is the most frequent target for inactivating mutations (46 occasions). However, such mutations in *rpoS* have been associated with long term storage condition and were not further considered (Bleibtreu *et al.*, 2014). The other candidates with evidence of convergent reductive evolution included *pepT* (14 occasions), *ynfK* (8 occasions) and *puuD* (7 occasions). *pepT* encodes for a tripeptidase which is highly expressed during biofilm formation, and is probably under the control of quorum-sensing (Prigent-combaret *et al.*, 1999). *ynfK* encodes for a putative dethiobiotin synthetase, which catalyses the penultimate step in biotin biosynthesis. However, I did not observe any inactivation mutations in the paralogue *bioD*, implying that the biotin biosynthesis pathway is more likely to be down regulated rather than inactivated. In *E. coli*/*Shigella*, biotin acts as the key co-factor for the acyl-CoA carboxylases (ACCs) to convert acetyl-CoA into fatty acid. Upon invading the host cell, *Shigella* imports fatty acids from the host cytosol. Therefore, the depletion in biotin may direct *Shigella*'s acetyl-CoA pool away from fatty acid production and conserve it for energy generation via the conversion of acetyl-CoA into acetate (Kentner *et al.*, 2014). Alternatively, *puuD* encodes for gamma-glutamyl-gamma-aminobutyrate hydrolase, an intermediate enzyme responsible for the conversion of putrescine into GABA. Mutations in *puuD*, together with pre-existing *puuE* pseudogene, can preferentially reroute putrescine to the production of spermidine, which scavenges free radicals and heightens resistance to oxidative stress (Barbagallo *et al.*, 2011). These examples show that though the aforementioned mutations are not fixed in the population, their multiple independent occurrences may

provide a transient fitness advantage for enhancing the pathogen's progress within a human host.

4.3.6 Substitution mutation analysis

Mutation analysis found that *rpoS* accumulated the highest number of substitutions since the emergence of the CenAsiaIII clade, but this likely is a laboratory artefact due to long-term storage in nutrient-deprived stab culture (Bleibtreu *et al.*, 2014). Additional genes with a high load of non-synonymous mutations may be under selection; these are summarized for this population in Table 4.3. Half of the candidate genes under selection are involved in response to antimicrobials (*gyrA* and *parC*), and cellular oxidation-reduction processes (*mreB*, *putA*, *ubiH*, *ydbK,* and *hyfB*). Several of these genes (*mreB*, *gyrA*, *ubiH*, and *hyf*) were previously identified as undergoing diversifying selection in the Vietnamese indigenous Global III *S. sonnei* population (Holt *et al.*, 2013), underlying that adaptation to oxidative stress is a common evolution trajectory of the pathogen. Notably, *ynfK*, which is commonly inactivated, is also likely under selection, demonstrating that partial cessation of biotin biosynthesis is favoured by *S. sonnei*.

Gene	No. of nonsynonymous mutations	No. of synonymous mutations	Adjusted dN/dS	Product and function
mreB	10	0	inf	Cytoskeleton protein, regulation of cell shape
gyrA	8	0	inf	DNA gyrase subunit A, response to antibiotic
putA	8	1	3.557	Proline dehydrogenase, proline to glutamate catabolism; oxidation – reduction process
sigA	7	0	inf	Serine protease, putative virulence factor
ygjU	6	0	inf	Serine/threonine symporter
ylbE	5	0	inf	Hypothetical protein
ynfK	5	0	inf	Dethiobiotin synthetase, biotin biosynthesis
ubiH	5	0	inf	Ubiquinone biosynthesis process, response to oxidative stress
parC	5	0	inf	DNA topoisomerase IV subunit A, response to antibiotic
yahG	5	1	2.22	Membrane
yieO	5	1	2.223	Putative transport protein MFS
SSON_1290	4	1	1.778	Hypothetical protein
ydbK	4	1	1.778	Pyruvate flavodoxin oxidoreductase, response to oxidative stress
hyfB	4	1	1.778	Hydrogenase 4, component B; oxidation – reduction process

Table 4.3. Genes likely under selection in CenAsiaIII *Shigella sonnei* population.

4.3.7 *Shigella sonnei* experimental evolution

Our recent diarrhoeal surveillance study performed in Ho Chi Minh city, Vietnam during 2014-2016 found that CenAsiaIII FQ resistant *S. sonnei* has effectively and gradually displaced the FQ susceptible indigenous population (Global 3 clade) (Duong *et al.*, 2018). I conducted a laboratory evolution experiment to examine differences in mutation accumulation between these two distinct lineages. The results of this experiment are summarized in Table 4.4. For the ciprofloxacin susceptible strain (cipS-VN; Global 3 clade), exposure to nalidixic acid and ciprofloxacin for nearly 450 generations increased the MIC against nalidixic acid (>256 μg/mL) and ciprofloxacin. The endpoint MICs for ciprofloxacin were variable among the various experimental conditions and replicates, but these did not reach the defined resistance breakpoint (1μg/mL). However, I did observe differing mutations between experimental conditions. Exposure to nalidixic acid consistently led to disruptive mutations in *emrR*, encoding for a transcriptional repressor. *emrR* mutations result in overexpression of the EmrAB pump, which heightens resistance against nalidixic acid (Lomovskaya, Lewis and Matin, 1995). Other explanatory mutations also included a SNP in *parE* and disruptive mutation in *acrR*. In contrast, *parE* appeared to be the frequent mutational target under ciprofloxacin pressure, with either a SNP or a deletion stably inherited in both replicate populations. It was apparent that *cysE* was consistently mutated across all culture conditions for cipS-VN but not for cipR-VN, with different mutations being stably maintained in a single population in two instances. *cysE* encodes for a serine acetyltransferase catalysing the conversion of L-serine to the quorum sensing signal O-acetyl-L-serine (OAS). Previous studies have reported that *cysE* mutants exhibit an increased rate of biofilm formation *in vitro* by suppressing the

production of OAS (Sturgill *et al.*, 2004). Additionally, minimal medium, as implemented here, is known to induce biofilm formation in *E. coli* (Naves *et al.*, 2008). Therefore, consistent *cysE* mutations could be a result of minimal medium culture. I also observed multiple independent mutations within the gene encoding the transcriptional repressor, *metJ*, which negates its regulatory effect and switches on the methionine biosynthesis pathway (Marincs *et al.*, 2006).

In addition, exposure of cipS-VN to ciprofloxacin, but not nalidixic acid, triggered a stable and specific genomic deletion (*treF-yhjA-yhi-hde-slp-iucABCD-iutA*). This region spans genes encoding for cytochrome C peroxidase (*yhjA*), acid stress resistance (*yhi* and *hde* loci) and aerobactin biosynthesis (*iuc* locus and *iutA*). In *S. sonnei*, this region is flanked by an IS4 element, which may explain the mechanism of genomic deletion. It has been recently proposed that bactericidal antimicrobials, including FQs, induce cell death via the production of reactive oxygen species (ROS) (Van Acker and Coenye, 2017). The common penultimate step in the production of ROS is the stimulation of the Fenton reaction, which recruits hydrogen peroxide and ferrous ions to produce hydroxyl radicals. These radicals can damage cellular macromolecules, specifically oxidizing the guanine nucleotide pool, and eventually leading to cell death. Consequently, high intracellular iron facilitates rapid FQ-mediated killing, and my results suggest that *S. sonnei* circumvents this by lowering the intracellular iron level via the obliteration of the siderophore aerobactin.

Condition	MIC-CIP (μg/mL)	MIC-NAL (μg/mL)	Resistance mutations	Other mutations (*)	Genomic deletion
Ciprofloxacin sensitive *S. sonnei* (Global III clade)					
D0	0.25	64 (sat)			
M9-A	0.19	64 (sat)		*cysE* (P252S)	
M9-B	0.25	64 (sat)		*cysE* (P252T); *clpA* (T374I); *metJ* (D76N, A61E); *rpoS* (Q161P)	
NAL-A	0.5	> 256	***emrR*** (fs); ***parE*** (P571S)	*cysE* (A33E); *cysE* (S249R); *cysB* (P145L)	
NAL-B	0.38	> 256	*acrR* (fs); *emrR* (fs)	*cysE* (Q228K); *dinG* (fs, 2 a.a deletion); *metJ* (R78P, A13T)	
CIP-A	0.38	> 256	***parE*** (S458L)	*cysE* (fs); *cysE* (P252T)	*treF-yhjA-yhi-hde-slp-iucABCD-iutA*
CIP-B	0.75	> 256	***parE*** (2 a.a insertion)	***cysE*** (M256I)	*treF-yhjA-yhi-hde-slp-iucABCD-iutA*
Ciprofloxacin resistant *S. sonnei* (Central Asia III clade)					
D0	8	>256			
M9-A	8	NA		***rpoS*** (R277L); ***iucA*** (stop codon)	
M9-B	6	NA		***rpoS*** (R301H); *iucD* (L113P); *ycjC* (P47S); SSON_2732 (2 a.a insertion)	
NAL-A	12	NA	*emrR* (fs)	SSON_1790 (stop codon); ***hemX*** (fs)	
NAL-B	8	NA	***emrR*** (I27 deletion)	**SSON_2588** (A93T)	
CIP-A	16 (sat)	NA	***rob*** (G245R)	eutQ (fs)	
CIP-B	16 (sat)	NA	*rob* (Q194K) *emrR* (fs)	SSON_1804 (N200K)	

Table 4.4 Results on antimicrobial susceptibility and mutation analysis of the *Shigella sonnei* experimental evolution. D0: initial condition on day 0. In all conditions, selected strains were cultured in minimal M9 medium supplemented with glucose and niacin (25 μg/mL) with the addition of: (1) M9-A/B: no antimicrobial, (2) NAL-A/B: the same concentration of nalidixic acid (256 μg/mL). (3) CIP-A/B: ciprofloxacin in concentrations of half the corresponding MIC of respective strain (0.125 μg/mL for cipS-VN and 4 μg/mL for cipR-VN). Mutations in bold indicate ones fixed in the examined population. Fs: frameshift mutation.

Exposure of cipR-VN to nalidixic acid similarly led to the generation of *emrR* mutants, which in one case caused an elevated MIC to ciprofloxacin. Long-term exposure of cipR-VN to ciprofloxacin doubled the MIC to ≥16μg/mL, and my analyses attributed this phenomenon to a mutation in *rob*. This gene encodes for the transcription activator Rob, a protein composed of two functional domains. The N-terminal domain shares substantial amino acid homology to SoxS (in response to oxidative stress) and MarA (in response to certain antimicrobials), indicating that Rob's regulon overlaps with those defined for these two common regulators (Chubiz, Glekas and Rao, 2012). It has been proposed that the C-terminal domain represses Rob's activity by sequestration, impeding access to transcriptional machinery (Griffith *et al.*, 2009). In my experiment, both mutations observed in *rob* fall into its C-terminal coding region, potentially changing its conformation and releasing more active Rob. Overexpression of Rob is linked to resistance to multiple antimicrobials, including quinolones (Ariza *et al.*, 1995), owing to its increased activation of multidrug efflux pumps (directly or indirectly via *marRAB*) and decreased expression of the outer membrane porin *ompF* (via upregulated *micF*) (Bennik *et al.*, 2000; Duval V. and Lister, 2014). In the absence of antimicrobials, cultured cipR-VN maintained fixed mutations in *rpoS*, both of which fell into the defined DNA binding residues of sigma38 ($E\sigma^S$)'s domains 4.1 (R277L) and 4.2 (R301H) (Campbell *et al.*, 2002). This region mediates $E\sigma^S$ binding to the -35 promoter region of regulated genes. However, most promoters recognized by $E\sigma^S$ show poor conservation in the -35 element, and mutagenesis of region 4 shows an insubstantial effect on promoter binding strength and specificity (Checroun *et al.*, 2004; Landini *et al.*, 2014). Therefore, the influence of observed mutations on *S. sonnei*'s *rpoS* regulon is hard to foresee, probably inclined

toward the degree of activation in this network under stress but not alternation in its components.

4.4 Discussion

This study is among the first attempt to utilize whole genome sequencing to examine the emergence and global spread of ciprofloxacin resistant *S. sonnei*. The extensive data show that all sequenced extant ciprofloxacin resistant *S. sonnei*, though sourced from disparate geographical locations, arose from a single major clone in early 2007. My analysis proposed that South Asia is the most likely reservoir for its origin and spread. These support previous hypotheses suggesting that ciprofloxacin resistant *S. sonnei* in industrialised countries is being imported from South Asia (Bowen *et al.*, 2015; De Lappe *et al.*, 2015). A recent estimation of worldwide antimicrobial usage reported that India was the largest consumer of antimicrobials in 2010 (Van Boeckel *et al.*, 2014). Additionally, the FQs are ranked as the most common antimicrobial prescribed for acute enteric diseases in India and Bangladesh (Fahad *et al.*, 2010; Kotwani, Chaudhury and Holloway, 2012). It is noticeable that the second and third mutations (*parC*-S80I, *gyrA*-D87G) were acquired rapidly in 2000s, and this adaptation is probably due to the sharply increased use of FQs during this period, when it became routinely prescribed for shigellosis (WHO, 2005; Laxminarayan and Chaudhury, 2016). The intensive use of FQs in a region where there are foci of high population density and inconsistent access to sanitation is likely to have contributed to emergence of ciprofloxacin resistant enteric bacteria, such as *S. sonnei* and *Salmonella* Typhi, on the Indian subcontinent (Wong *et*

al., 2015). Global dissemination of these organisms is likely facilitated by the volume of travel between these regions and other areas of the world.

The new data highlight the limitations of current typing protocols for tracking *S. sonnei*. It had been previously observed that some of the ciprofloxacin resistant *S. sonnei* isolates in this study (originating from Bhutan and Ireland) shared a similar *Xba*I Pulse Field Gel Electrophoresis (PFGE) pattern (Ruekit *et al.*, 2014; De Lappe *et al.*, 2015). This pulsotype has been observed previously in India and Bangladesh (Talukder *et al.*, 2006; Pazhani *et al.*, 2008; Ghosh *et al.*, 2011; Nandy *et al.*, 2011; Ud-Din *et al.*, 2013), as well as in Canada (Gaudreau *et al.*, 2011), Belgium (Vrints *et al.*, 2009) and Japan (Uchimura, Kishida and Koiwai, 2001), where the association with ciprofloxacin resistance was inconsistent. However, PFGE in this context did not offer sufficient granularity to link all of the isolates or provide sufficient resolution into the regional evolution of *S. sonnei*. The phylogenetic analyses in this chapter show that this pulsotype is associated with a phylogenetic lineage, supporting the notion that this pulsotype actually represents a widespread and pervasive subclade of CenAsia III.

I showed evidence of FQ resistant *S. sonnei*'s sustained circulations in Southeast Asia and Europe, but its degree of propagation is likely to be underestimated in regions not representative in the analysis such as the East Asia, Africa and America. Indeed, the same clone was confirmed to cause a recent outbreak in California in 2015 (Kozyreva *et al.*, 2016). Besides, FQ resistant *S. sonnei* harbouring the identical QRDR mutations is prevalent among the MSM community in Taiwan (Chiou *et al.*, 2016). These suggest that

the organism is spreading rapidly worldwide and poses a significant threat to public health. Particularly, its transmission within the MSM network should be closely monitored, especially in the wake of increased shigellosis incidence and antimicrobial resistance within this high-risk group (Baker *et al.*, 2015; Mohan *et al.*, 2017). Though I report here the major resistance clone accumulated triple QRDR mutations, I could not exclude the existence of other mechanisms. A year-long surveillance of *S. sonnei* in Wales and England identified 7 FQ resistant strains with the combinatorial mutations *gyrA*-S83L, *parC*-S80I and *gyrA*-D87N, which is identical to a single resistant outlier in my collection (Sadouki *et al.*, 2017). Resistance due to the synergy of plasmid mediated *qnrB* and *gyrA* mutations have been noted in India as well (Das, Natarajan and Mandal, 2016). The frequent gains of resistance determinants to other first line treatments, ESBL and *mphA*, further exacerbate the situation, especially in Southeast Asia where local Enterobacteriaceae gene pool is vast and filled with resistance genes.

My genomic examination pointed to the prevalence of reductive evolution mechanisms in CenAsiaIII *S. sonnei*, even under the prism of thirty years. Though it is not possible to fully catalogue pseudogenes generated via IS element disruptions, I observed some noticeable evolutionary trend. These include the disruption of several metabolic functions, specifically the fermentative processes of alternate carbon sources, as well as the cessation of cellular appendages such as flagella and pili. Though *S. sonnei* is conventionally described as non-motile, flagella production has been observed albeit with much lower magnitude compared to other Enterobacteriaceae (Girón, 1995). I showed that by shedding the structural gene *fliC* and its transcription unit (*fliAZY*) entirely,

CenAsiaIII *S. sonnei* evolved to be totally independent of motility in its lifestyle. In *E. coli*, loss of *fliA* led to better survival during stress, and this is due to the increased energy conservation and membrane physical stabilization, which resulted from the loss of motor protein and flagellum apparatus (Fontaine *et al.*, 2008). A previous study also showed that non-flagellated *P. putida* possesses higher endurance under oxidative stress via releasing more energy and NADPH, which is otherwise coupled to the flagellar machinery (Martínez-García *et al.*, 2014). Furthermore, similar to the loss of *fliA*, my analysis pointed to a selection for better adaptation under oxidative stress. Truncation in *yeaK* increases translational errors under excessive serine/threonine condition, but tolerable mistranslation was shown to actually increase fitness in *E. coli* subjected to oxidative stress (Fan *et al.*, 2015; Liu *et al.*, 2015). Adaptation under oxidative stress is further supported by the finding that most of the genes identified to be under selection are linked to oxidation-reduction processes. My conjecture is that the high pressure of oxidative stress may come from varied sources, but most likely due to the oxidative environment within macrophages or exposure to antimicrobials such as the FQs. Better survival during stress could benefit FQ resistant *S. sonnei* clone in its successful expansion into different geographical destinations. However, this interpretation is subjected to scrutiny since dN/dS rate is not designed for intraspecific rapidly evolving bacterial populations (Kryazhimskiy and Plotkin, 2008). Specifically, the nucleotide differences observed within a population do not represent fixed divergences but rather transient polymorphisms, which are not necessarily purged from the population due to the short evolution timeframe. Though dN/dS > 1 still indicates the presence of positive selection, this ratio does not increase monotonically with the selection coefficient.

Exposure to antimicrobial of class quinolone resulted in similar mutational profile for both FQ resistant and sensitive *S. sonnei* in spite of their different genetic backgrounds. In addition, I showed here for the first time that mutations in the transcriptional activator *rob* further produce resistance to FQ at very high MIC, prompting that mechanisms for elevated resistance is probable in this FQ resistant organism. It is intriguing to find that despite harbouring a *gyrA* mutation, continuous exposure of FQ susceptible *S. sonnei* to ciprofloxacin failed to generate specific resistance mutations in *gyrA* and *parC*. Previous in vitro studies also reported that in *E. coli* with single *gyrA* mutation background, mutations linked to drug efflux but not specific QRDR targets are preferentially selected (Kern *et al.*, 2000). This *in vitro* finding also corroborate with the scenario in which indigenous *S. sonnei* in Vietnam did not accumulate triple mutations to achieve FQ resistance phenotype even though the antimicrobial is heavily prescribed in this settting. This stands in contrast to the ease of obtaining FQ resistance via QRDR mutations in other bacteria, such as *Staphyloccocus aureus*. This species presumably has much bigger population size than *S. sonnei*, and FQ resistance only requires double mutations in *gyrA* and *grlA*, thus explaining its much higher occurrences across *S. aureus*'s phylogeny. FQ resistance attributed to QRDR mutations has been shown to be not associated with detrimental fitness cost in Enterobacteriaceae *in vitro*, even in the absence of FQ (Marcusson, Frimodt-Møller and Hughes, 2009; Baker *et al.*, 2013). Here I predict that, in the absence of quinolone pressure, FQ resistant *S. sonnei* showed mutational bias toward alternating the *rpoS* induced stress response. Since the experimental evolution only included limited number of replicates, future larger studies are necessary to confirm whether *rpoS* mutations are truly compensatory for the resistance phenotype.

5 Southeast Asian *Shigella* from a historical perspective

5.1 Abstract

Shigella is responsible for a substantial burden of diarrhoea in young children in developing countries, including several in Southeast Asia. Conventional disease surveillance for *Shigella* associated diarrhoea has relied on serotyping and low-resolution molecular typing, which fails to contextualize the evolutionary history of the genus. Here, I examined a collection of more than 1,800 *Shigella* whole genome sequences from bacteria isolated in four continental Southeast Asian countries (Thailand, Vietnam, Laos, and Cambodia) over the past three decades to characterise the evolution and transmission of both *S. flexneri* and *S. sonnei*. These data show that heterogeneity in population structure delineates the epidemiology of both species. The majority of these genetically distinct *Shigella* clones were likely introduced into the region in the 1980s-1990s. Within country, as well as trans-border, dissemination allowed widespread propagation of both species in the region. Unpredictably, the epidemiology of *S. sonnei* and major *S. flexneri* serotypes was characterized by frequent clonal replacement events. Such population alternations coincided with increasing resistance to antimicrobials, most notably co-trimoxazole in *S. flexneri* and the FQs in *S. sonnei*. This indicates that adaptation to antimicrobial pressure may have driven the recent evolutionary trajectory of *Shigella*. This work provides a new understanding of the evolution of shigellosis in Southeast Asia, revealing the population dynamics of its etiological agents in this endemic region.

5.2 Introduction

Shigella remains a common cause of dysentery in young children in developing countries, including several in Southeast Asia (SEA), such as Vietnam, Thailand, Laos and Cambodia (Bodhidatta *et al.*, 2002; von Seidlein *et al.*, 2006; Meng *et al.*, 2011; Thompson *et al.*, 2015). Regional epidemiological surveillance is important in characterizing the burden and distribution of different *Shigella* species and serotypes, but this approach does not provide an in-depth understanding of the pathogen's population structure and life history. The high resolution offered by WGS provides unprecedented granularity for studying organism circulation and evolution. A pioneering study utilizing WGS of contemporaneous *S. sonnei* detailed how the pathogen evolved and spread in South Vietnam following its introduction into the country in the 1980s (Holt *et al.*, 2013). This study also emphasized that antimicrobial resistance, either via mutations or acquisitions of mobile elements, is crucial in shaping the clonal expansion of *S. sonnei*. Nevertheless, much is unknown about the evolutionary history of another important aetiological agent, *S. flexneri*, as well as of the genus *Shigella* in other SEA countries.

This chapter gathered a vast collection of historical *Shigella* genome sequences in order to investigate the population structure, evolution and transmission of both *S. sonnei* and *S. flexneri* in continental Southeast Asia. I additionally examined how antimicrobial resistance impacts the population dynamics of the two species in this region.

5.3 Results

5.3.1 Southeast Asian *Shigella* in global context

This large-scale sequencing effort compiled a collection of 1,804 *Shigella* sequences across mainland SEA, including 803 *S. sonnei* and 1,001 *S. flexneri* isolates. These were retrieved from multiple collaborative institutions in the region, including: AFRIMS Thailand, N=1,239; Lao-Oxford-Mahosot Hospital-Wellcome Trust Research Unit (LOMWRU), N=43; Cambodia-Oxford Medical Research Unit (COMRU), N=19; Pasteur Institute in Cambodia, N=19; OUCRU Ho Chi Minh city, N=113. Figure 5.1 details the geographical distribution of these samples in the region, with Bangkok, central Thailand, and the south of Vietnam accounting for half of the collection (n=915). The majority of isolates from Northern and Southern Vietnam originated from the capital Hanoi and the urban Ho Chi Minh City (HCMC), respectively. Thus, these regions are used interchangeably throughout this chapter. *Shigella* serotyping was performed in each respective institution, but this data was not available for every isolate. Thus, I applied molecular serotyping (or geno-serotyping) to all *S. flexneri* strains based on the detection of specific combinations of the phage-borne serotype determining factors, *gtr* and *oac* variants. In line with previous surveillance efforts, *S. flexneri* 2a (n=431) was determined to be the predominant serotype in the region, followed by serotype 3a (n=157) (Bangtrakulnonth *et al.*, 2008; Vinh *et al.*, 2009).

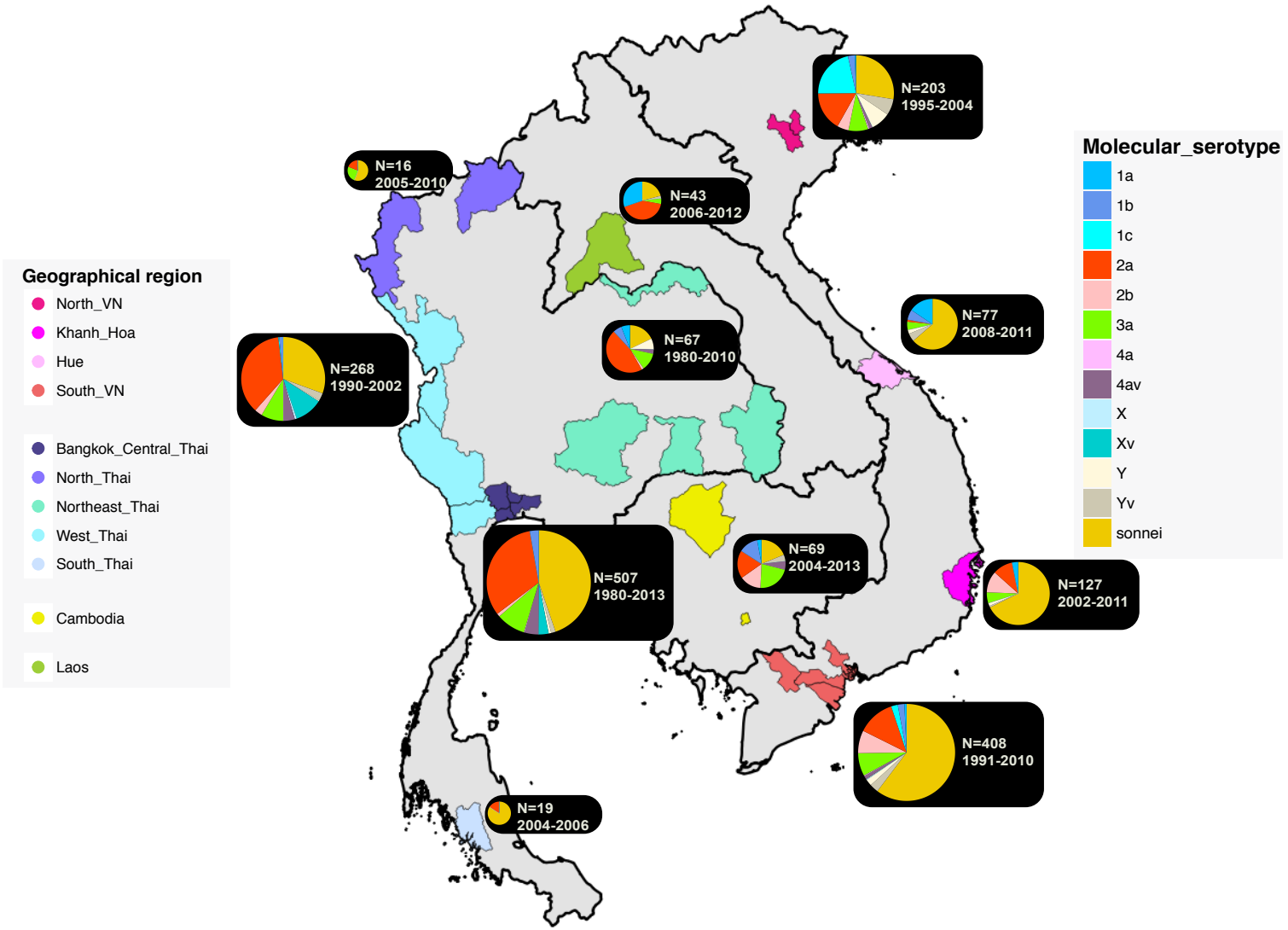

Geographical region
North_VN
Khanh_Hoa
Hue
South_VN
Bangkok_Central_Thai
North_Thai
Northeast_Thai
West_Thai
South_Thai
Cambodia
Laos
Molecular_serotype
1a
1b
1c
2a
2b
3a
4a
4av
X
Xv
Y
Yv
sonnei
N=203
1995-2004
N=16
2005-2010
N=43
2006-2012
N=77
2008-2011
N=268
1990-2002
N=67
1980-2010
N=69
2004-2013
N=507
1980-2013
N=127
2002-2011
N=408
1991-2010
N=19
2004-2006

Figure 5.1 Overview of Southeast Asian *Shigella* genome sequences used in this study. The figure displays a map of four Southeast Asian countries where the *Shigella* samples originated (Thailand, Cambodia, Laos, and Vietnam). Cities/provinces with collected *Shigella* sequences are highlighted with respect to the country's geographical divisions, except for Laos (where the capital, Vientiane, is highlighted to represent Laotian isolates). These regions include Bangkok_ Central_Thai (Bangkok, Nonthaburi, Samut Sakhon, Nakhon Pathom); North_Thai (Chiang Rai, Mae Hong Son); Northeast_Thai (Nong Khai, Ubon Ratchathani, Surin, Nakhon Ratchasima); West_Thai (Tak, Kanchanaburi, Ratchaburi); South_Thai (Trang); Cambodia (Phnom Penh, Siem Reap); North_VN (Hanoi); Hue; Khanh_Hoa; South_VN (Ho Chi Minh City, Dong Thap, Tien Giang, Ben Tre). The pie charts show the proportion of all *Shigella* molecular serotypes, as assessed by genomic analysis (see section 2.4.2), identified from the corresponding region (see Legend). The size of pie charts is proportional to the total number of isolates sequenced in the region. This is accompanied by a year range indicating the duration of sampling.

In order to provide a phylogenetic context for SEA *S. flexneri*, I combined my dataset with more than 300 public sequences from the species global diversity study to construct a maximum likelihood phylogeny (Connor *et al.*, 2015) (Fig 5.2). This analysis showed that organisms isolated in the past three decades were positioned in all previously defined lineages (or phylogenetic groups, PGs), with the exception of Lineages 5 and 6. One novel lineage, consisting exclusively of isolates from SEA, was most closely related to Lineage 7 and was defined as Lineage 8 in this study. The over-representation of *S. flexneri* Lineage 3 in the region is reasonable since it is mainly composed of 2a, the most prevalent serotype. To enhance phylogenetic resolution, sublineage division was conducted for Lineages 2 and 3. Clear geographical clustering was observed for Lineage 2, with 2.1 and 2.2 composed only of Asian samples and 2p mainly endemic in Africa. Lineage 3 could be subdivided into three sublineages (3Xv, 3.1, 3.2), with Sublineages 3.1 and 3.2 appearing to be of greatest epidemiological significance.

The global phylogenetic structure of *S. sonnei* confirmed that organisms sampled across SEA mainly belonged to Lineage III (Fig 5.3). These organisms were closely related, however the whole genome phylogeny allowed the delineation of this lineage into three clusters. Two distinct clades were nested within the internationally disseminated Global 3, and thus were named Global 3.1 and 3.2 (Holt *et al.*, 2012). Another clade, termed early-SEA, represents a clonal expansion that was exclusively found only in SEA, and was most closely related to the South American expansion of *S. sonnei* in the 1950s. A group of over 30 SEA *S. sonnei* were clustered within Lineage I, which is atypical given that this lineage has been previously associated with historical European *S. sonnei* only.

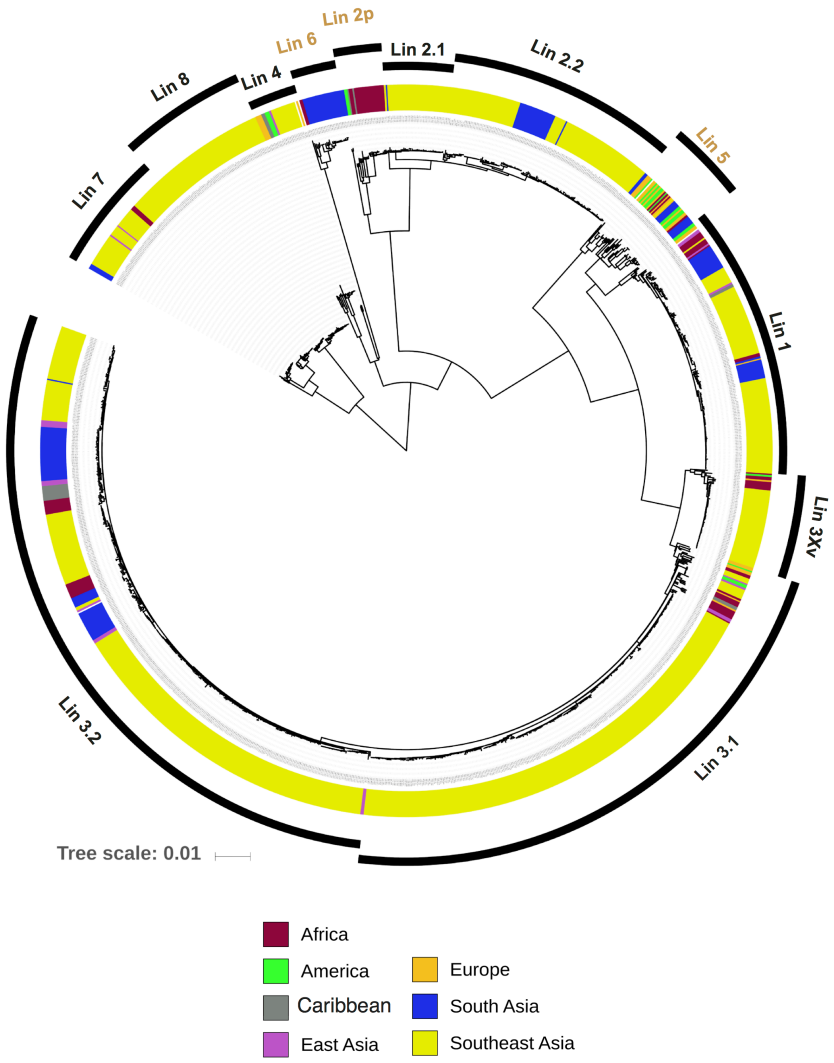


Figure 5.2 The global phylogenetic structure of *Shigella flexneri*. The figure displays the maximum likelihood phylogeny of 1,323 *S. flexneri* isolates, compiled from isolates sequenced for this study and those in a previously published global *S. flexneri* diversity study (Connor *et al.*, 2015). Different lineages were identified based on Bayesian Analysis of Population Structure (BAPS) hierarchical clustering and were named in accordance to the nomenclatures defined from the aforementioned study (Lineages 1 to 7), with Lineage 8 newly defined in the current study. The tree is rooted using Lineages 7 and 8 (section 2.4.3). The colours of the inner ring indicate the isolate's geographical source (see Legend) while the outer ring defines the lineage clustering. Lineages which do not contain substantial numbers of Southeast Asian sequences are labelled in orange and excluded from further analysis. The horizontal scale bar shows the number of nucleotide substitutions per site.

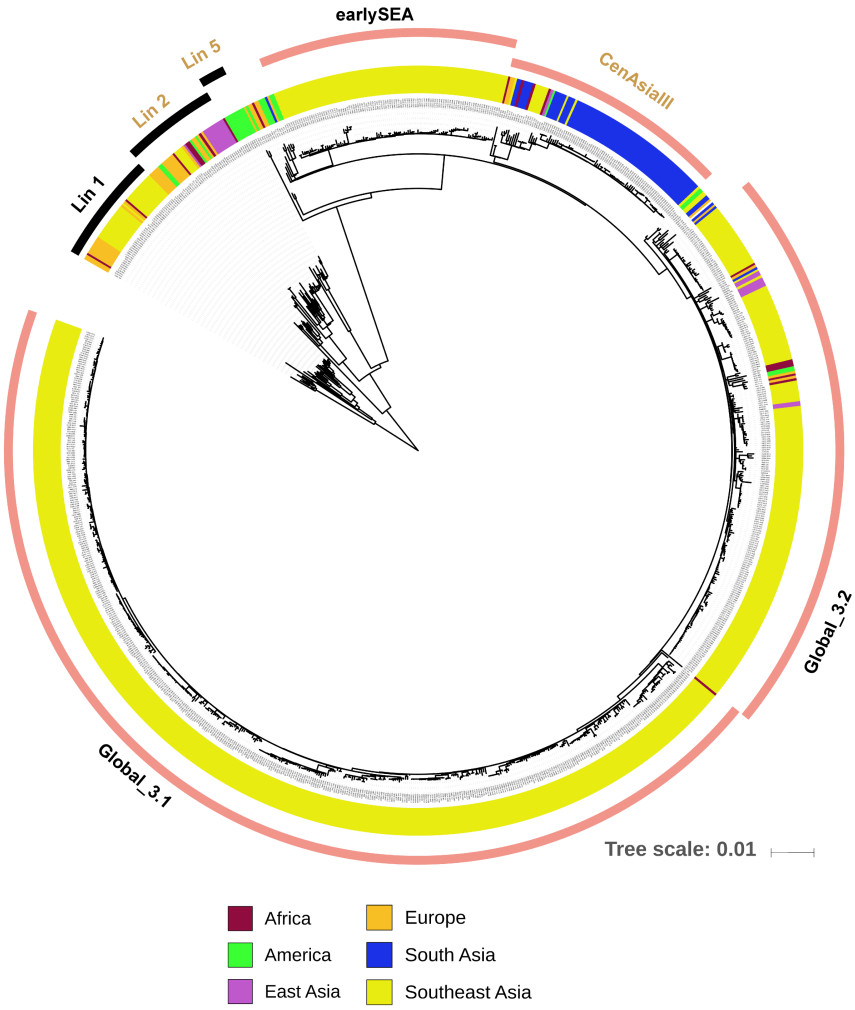


Figure 5.3 The global phylogenetic structure of *Shigella sonnei*. The figure displays the maximum likelihood phylogeny of 993 *S. sonnei* isolates, compiled from those sequenced for this study and those published previously for *S. sonnei* global phylogenetic study (Holt *et al.*, 2012), *S. sonnei* local evolution in Vietnam (Holt *et al.*, 2013), and *S. sonnei* local evolution in Bhutan (Chung The *et al.*, 2015). The isolates are grouped into lineages defined previously in the global diversity study, and the tree is rooted using Lineage 1 as an outgroup. The colours of the inner ring indicate the isolate's geographical source (see Legend) while the outer ring defines the lineage clustering (with the sublineages comprising Lineage 3 coloured in coral). Lineages which do not contain substantial numbers of Southeast Asian sequences are labelled in orange and excluded from downstream analysis. The horizontal scale bar indicates the number of nucleotide substitutions per site.

5.3.2 Co-circulation of multiple *Shigella* lineages in Southeast Asia

The concurrent propagation of diverse *Shigella* populations within a geography complicates an understanding of the genus's evolutionary history. This co-circulation of different *Shigella* is apparent when the epidemiological contribution of each lineage is plotted temporally within each nation, as shown in Figure 5.4. Although the overall sampling period spans more than three decades, temporal discrepancies between regions are evident, with sequences from Thai *Shigella* dominating datasets in 1980-1995 and Vietnamese *Shigella* prevailing in the period from 1995 onward. Sequences from Laos and Cambodia were only available from the mid 2000s. I additionally observed the diminishing isolation rate of *S. flexneri* in Thailand from the mid-1990s, which mirrors the trend reported by recent human shigellosis surveillance in Thailand (Bangtrakulnonth *et al.*, 2008). Similarly, the soar of Vietnamese *S. sonnei* in this dataset since the mid-2000s also matches the epidemiological observation in Southern Vietnam (Vinh *et al.*, 2009). These data reflect species replacement event occurring in these two countries. In order to investigate the evolutionary history of SEA *Shigella*, including *S. flexneri* and *S. sonnei*, within each lineage, I constructed individual maximum likelihood phylogenies for each examined clade. These offer enhanced resolution of phylogenetic relationships at nodes near the tips of the tree, which are rarely fully resolved in the context of the species' global population structure. All lineages, except for *S. flexneri* Lineage 4, displayed a high correlation between sampling dates and the phylogeny's root-to-tip divergence (R^2 ranging from 0.5 to 0.95; linear regression), underlying the validity of the molecular clock hypothesis in the evolution of this pathogen (Fig 5.5 and 5.6). However, this interpretation needs to take into account the confounding effect caused by unequal sampling frequency in each lineage-specific phylogeny.

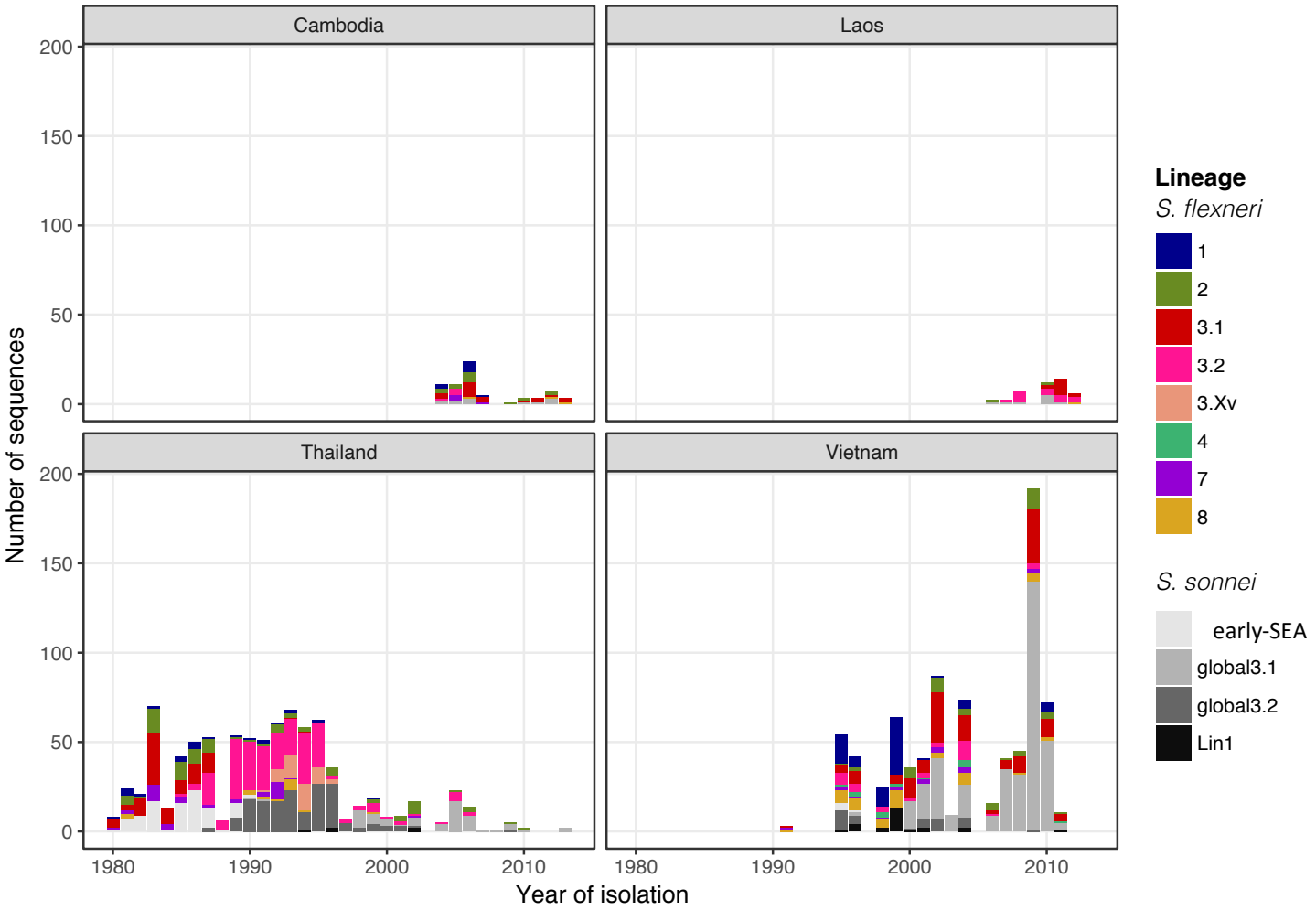


Figure 5.4 Summary of *Shigella* sampling timeframe employed in this study. The number of sequences belonging to each defined *Shigella* lineage was plotted per calendar year for each SEA country (see Legend).

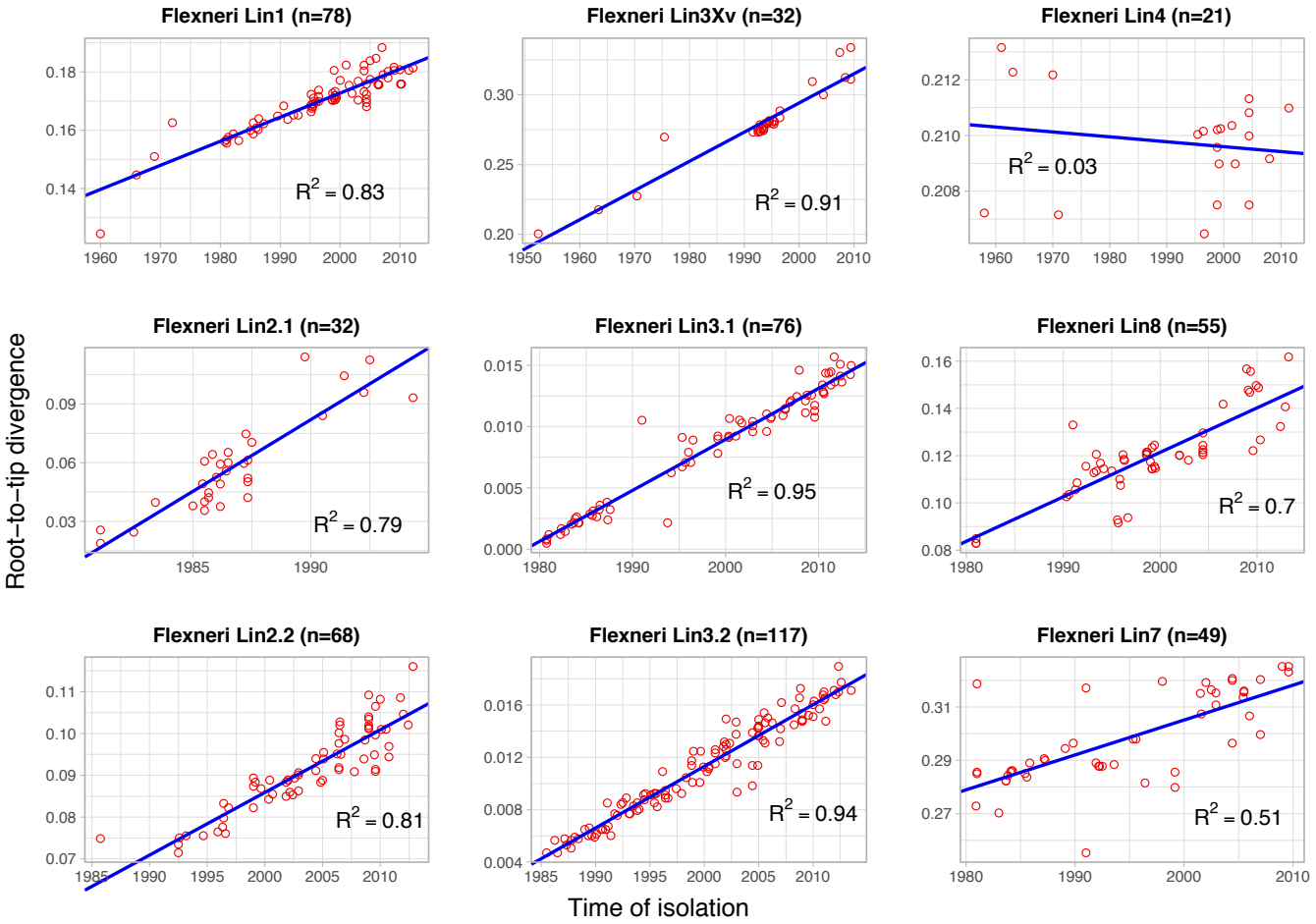


Figure 5.5 Temporal phylogenetic signal for *Shigella flexneri*. The figure shows the correlations between the time of isolation (in months and years; x-axis) and root-to-tip divergence (y-axis) for each *S. flexneri* lineage or sublineage. The linear regression line is coloured blue, and the coefficient of determination (R^2) is shown for each plot. Estimation of temporal signal was performed using TempEst v1.5.1, using the reference mapped SNP-based maximum likelihood phylogeny of each *S. flexneri* lineage or sublineage.

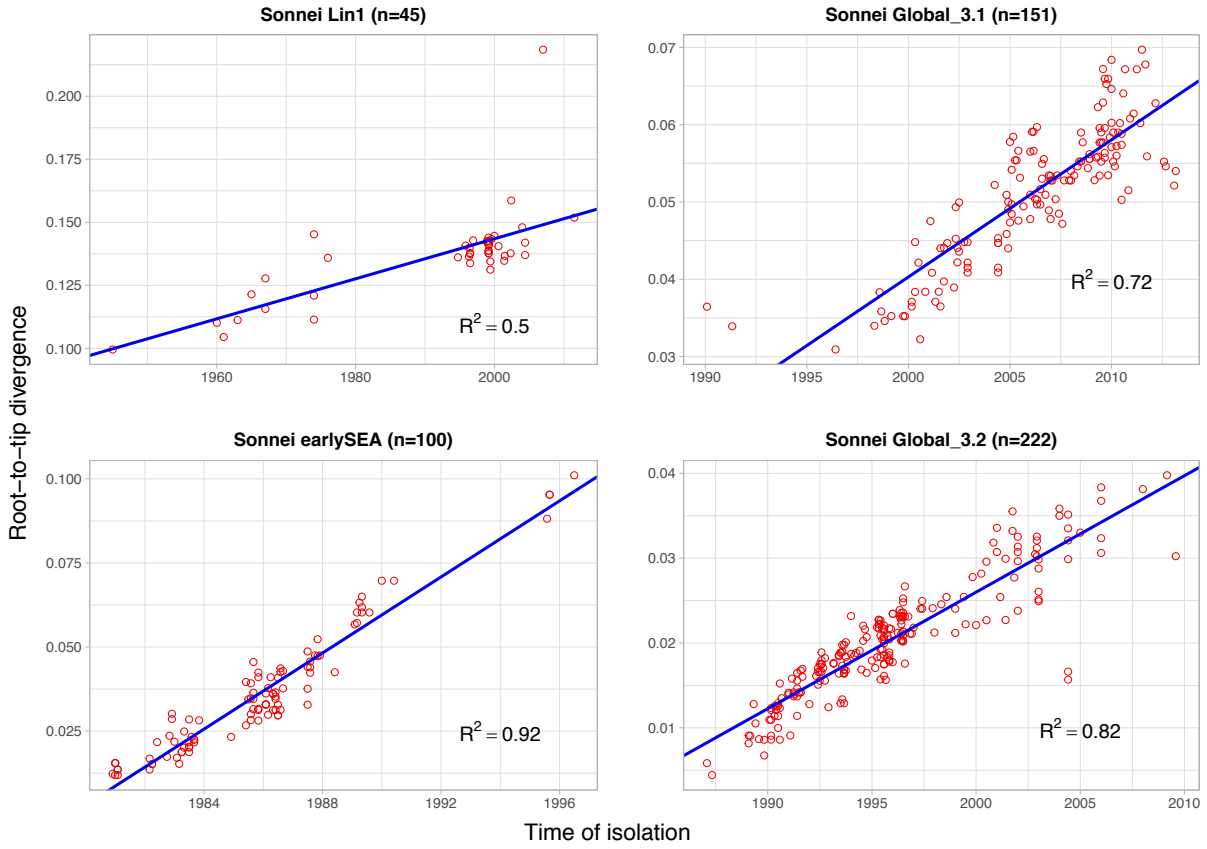


Figure 5.6 Temporal phylogenetic signal for *Shigella sonnei*. The figure shows the correlations between the time of isolation (in months and years; x-axis) and root-to-tip divergence (y-axis) for *S. sonnei* Lineage 1 and three sublineages of Lineage 3. The linear regression line is coloured blue, and the coefficient of determination (R^2) is shown for each plot. Estimation of temporal signal was performed by TempEst v1.5.1, using the reference mapped SNP-based maximum likelihood phylogeny of each *S. sonnei* lineage and sublineage.

Inspection of *Shigella* genomes at lineage level revealed that recombination is infrequent (Fig 5.7 & 5.8). These identified recombination elements were removed for further analyses. Bayesian phylogenetic inferences permit an estimation of the tMRCA of multiple SEA *Shigella* populations, which outlines their evolutionary histories in the region (Table 5.1). These were carried out for each lineage according to methods described in section 2.4.4. Figure 5.9 summarizes the timescale for the circulation of major *Shigella* lineages examined in this study. Despite their overarching genetic differences, the majority of these populations only emerged in the region during the last century, most commonly in the 1980s-1990s. Some lineages, such as the *S. sonnei* early-SEA clade and *S. flexneri* 3Xv, were composed of a single population exhibiting transient establishment in the region; due to spatial and temporal bias in sampling, their dissemination across the region could not be investigated. In contrast, other major *Shigella* lineages displayed a high degree of genetic diversity and a wide geographical distributions. Therefore, I focused on the epidemiologically important lineages (*S. flexneri* sublineages 3.1, 3.2 and *S. sonnei* Global 3.1 and 3.2) to examine the pathogen's mobility within SEA. These lineages were selected to quantify the degree of transitions between geographical states, utilizing stochastic mapping with even sampling. My results demonstrated that the two species shared a similar transmission pattern during their regional evolution. Firstly, within Thailand, *Shigella* was frequently sourced from Bangkok and the Central to Western, and Northeastern provinces, creating pan-regional *Shigella* populations (Fig 5.10). Secondly, I observed that *Shigella* were introduced from Central Thailand into Vietnam via major urban areas such as HCMC and Hanoi. Thirdly, extensive radiation from HCMC to other areas (Hue, Khanh Hoa, and Hanoi) characterized the *Shigella* transmission network within Vietnam.

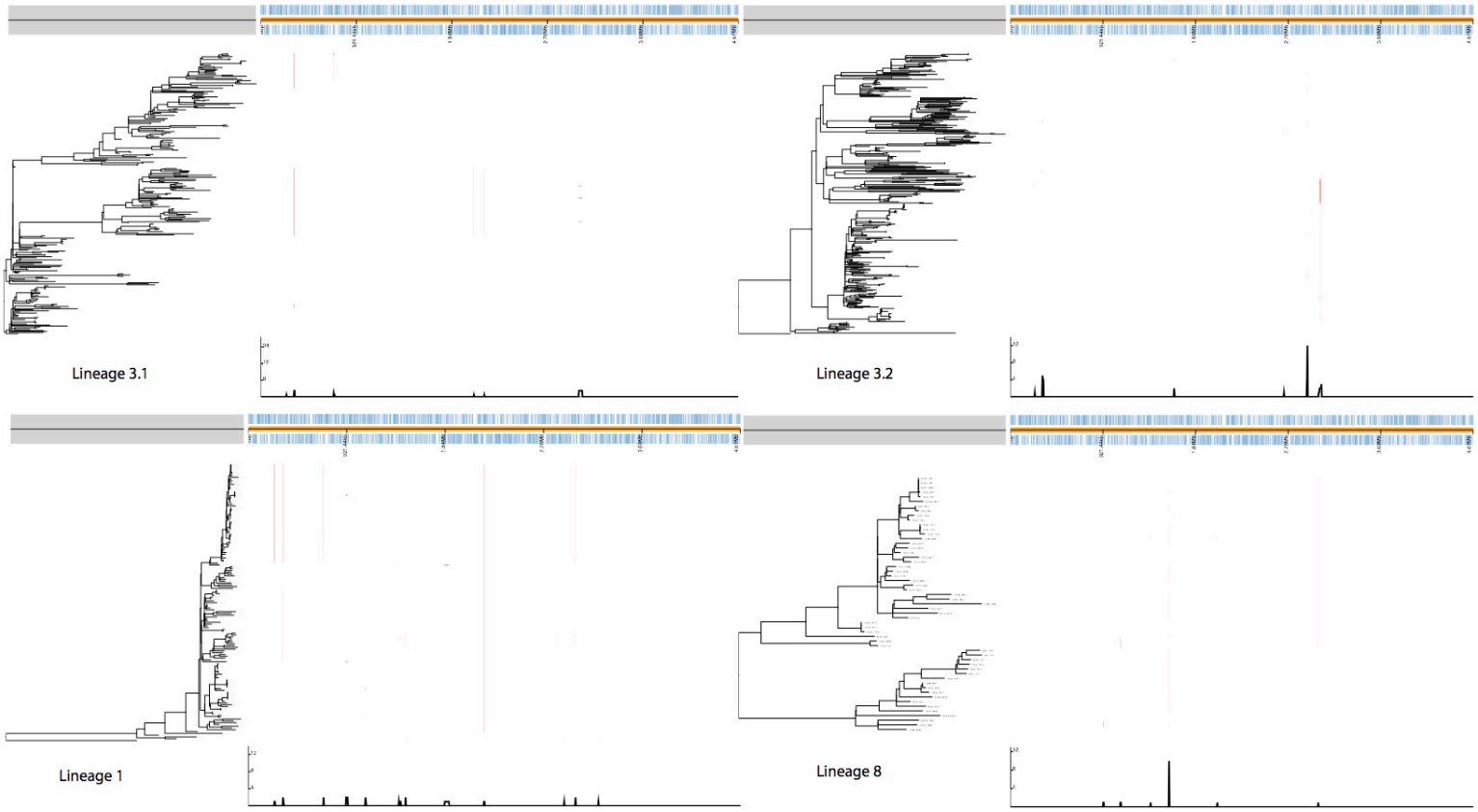


Figure 5.7 Recombination features within representative Southeast Asian *Shigella flexneri* lineages.

For each panel, the phylogeny on the left is constructed for each major lineage using maximum likelihood method. The top right bar displays the genomic coordinates of *S. flexneri* reference Sf2a 301, with coding sequences coloured in blue. Presumed recombination regions are detected within the genomes using Gubbins, which are highlighted in blue (in single taxon) or red (shared between many monophyletic taxa).

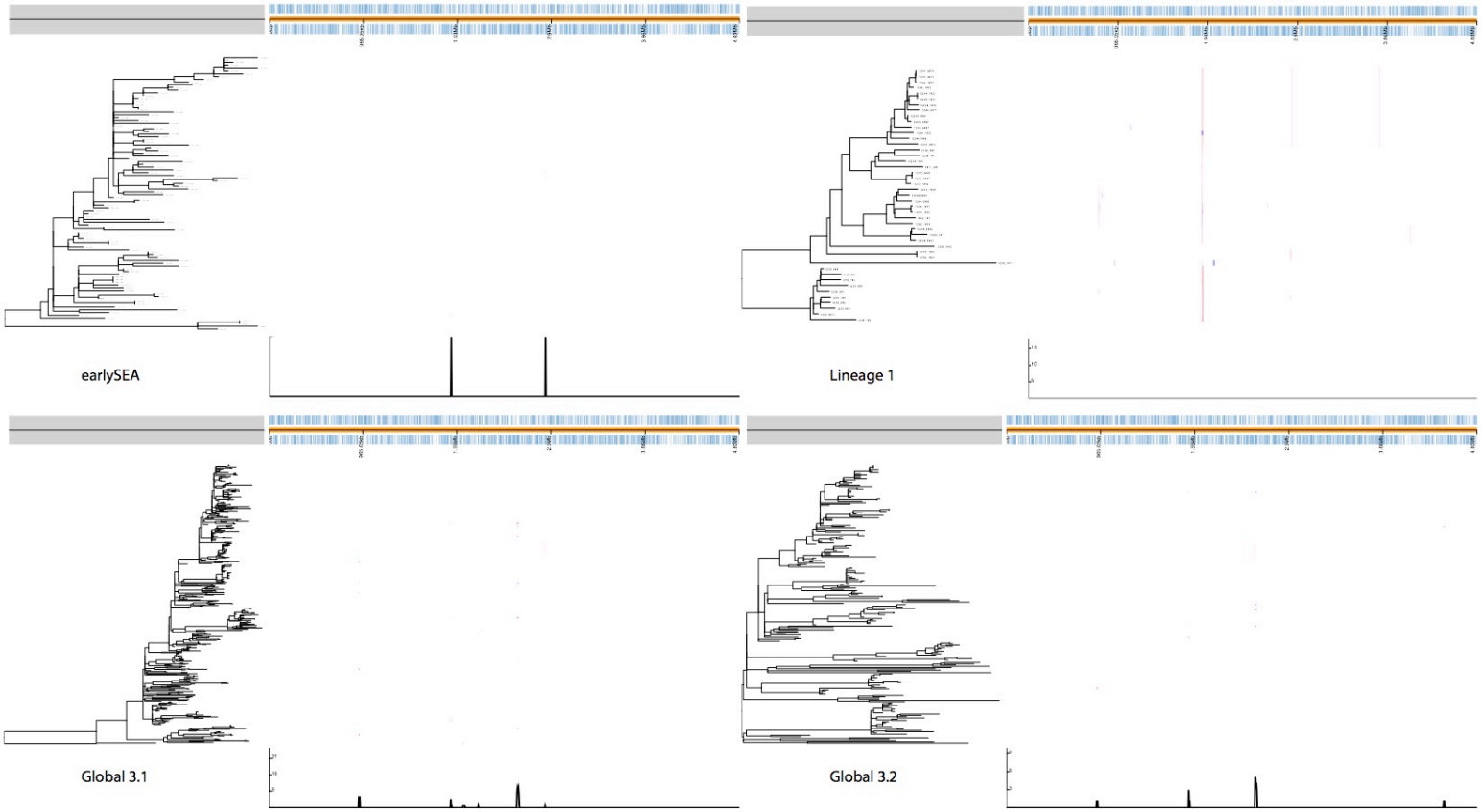


Figure 5.8 Recombination features within representative Southeast Asian *Shigella sonnei* lineages.

For each panel, the phylogeny on the left is constructed for each major lineage using maximum likelihood method. The top right bar displays the genomic coordinates of *S. sonnei* reference Ss046, with coding sequences coloured in blue. Presumed recombination regions are detected within the genomes using Gubbins, which are highlighted in blue (in single taxon) or red (shared between many monophyletic taxa).

Species	Lineage	tMRCA of SEA isolates	95% HPD-lower	95% HPD-upper	Substi-tution model	Clock model	Demogra-phic model	Substitution rate	95% HPD-lower	95% HPD-upper
Flexneri	1	1974.43	1972.36	1976.5	TVM	R-local	B-skyline	6.65E-07	6.10E-07	7.25E-07
	2.1	1978.79	1977.26	1980.57	K3P	R-local	B-skyline	7.58E-07	6.40E-07	8.70E-07
	2.2	1948.69	1942.15	1955.3	TVM	Strict	B-skyline	7.52E-07	6.70E-07	8.31E-07
	3Xv	1988.81	1987.79	1990.04	TVM	R-local	Constant	8.60E-07	7.50E-07	9.60E-07
	3.1	1978.66	1977.74	1979.62	TVM	R-local	B-skyline	7.98E-07	7.4E-07	8.56E-07
	3.2	1979.58	1977.11	1982.29	TVM	R-local	B-skyline	8.85E-07	8.02E-07	9.61E-07
	7	1400.09	1236.37	1589.93	TVM	Strict	Constant	1.40E-07	1.00E-07	1.78E-07
	8	1932.57	1915.11	1953.65	TVM	R-local	B-skyline	3.66E-07	2.78E-07	4.49E-07
Sonnei	Lin1	1869.93	1834.41	1911.2	TVM	R-local	Constant	3.84E-07	3.04E-07	4.63E-07
	earlySEA	1979.36	1978.62	1980.19	TVM	R-local	B-skyline	8.89E-07	7.63E-07	1.02E-06
	Global 3.1	1983.32	1980.75	1986.18	TVM	R-local	B-skyline	5.42E-07	4.81E-07	6.04E-07
	Global 3.2	1984.67	1984	1986.05	TVM	R-local	B-skyline	7.99E-07	7.25E-07	8.70E-07

Table 5.1 Summary of findings from Bayesian phylogenetic inference for major *Shigella* lineages. The table details the estimation of major parameters, including tMRCA of isolates in Southeast Asia and its 95% high posterior density (95% HPD), the substitution rate (in substitutions per site per year) and its 95% HPD. Marginal likelihood was calculated by both path sampling and stepping-stone sampling approaches, which ensures robust selection of the most appropriate model. R-local: Random local; B-skyline: Bayesian skyline.

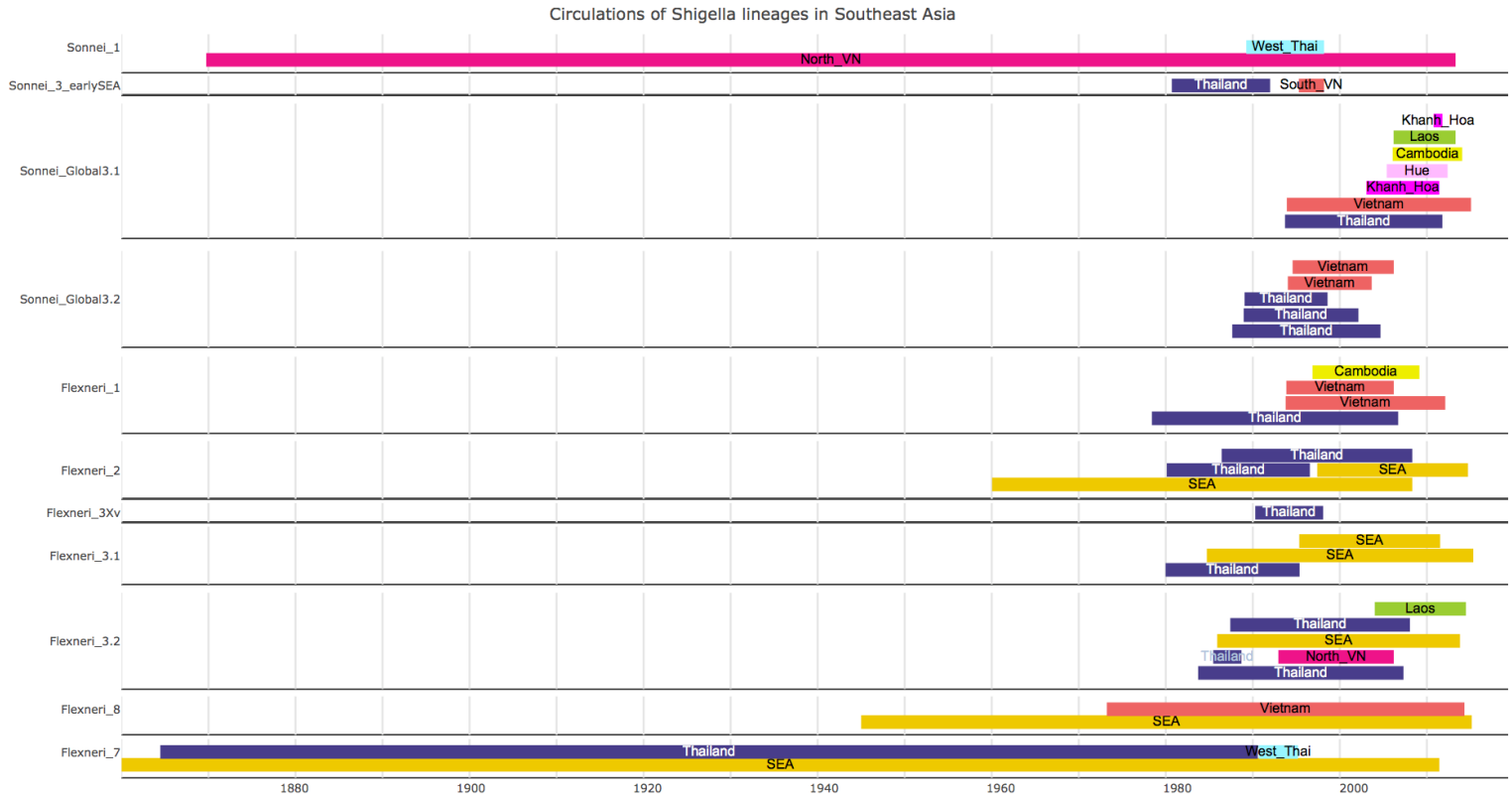


Figure 5.9 Circulations of different *Shigella* lineages in Southeast Asia. The timeline summarizes the circulations of multiple *Shigella* populations in different Southeast Asian regions. Each horizontal bar represents a population emergence or local expansion event, and starts with the tMRCA of the respective population in Southeast Asia (as estimated by Bayesian phylogenetic inference) and ends with the most recent isolate's sampling date. Each bar is coloured and named based on the distinct geographical region in which the *Shigella* population mainly circulates, with SEA denoting the concurrent presence of the population in more than one Southeast Asian country.

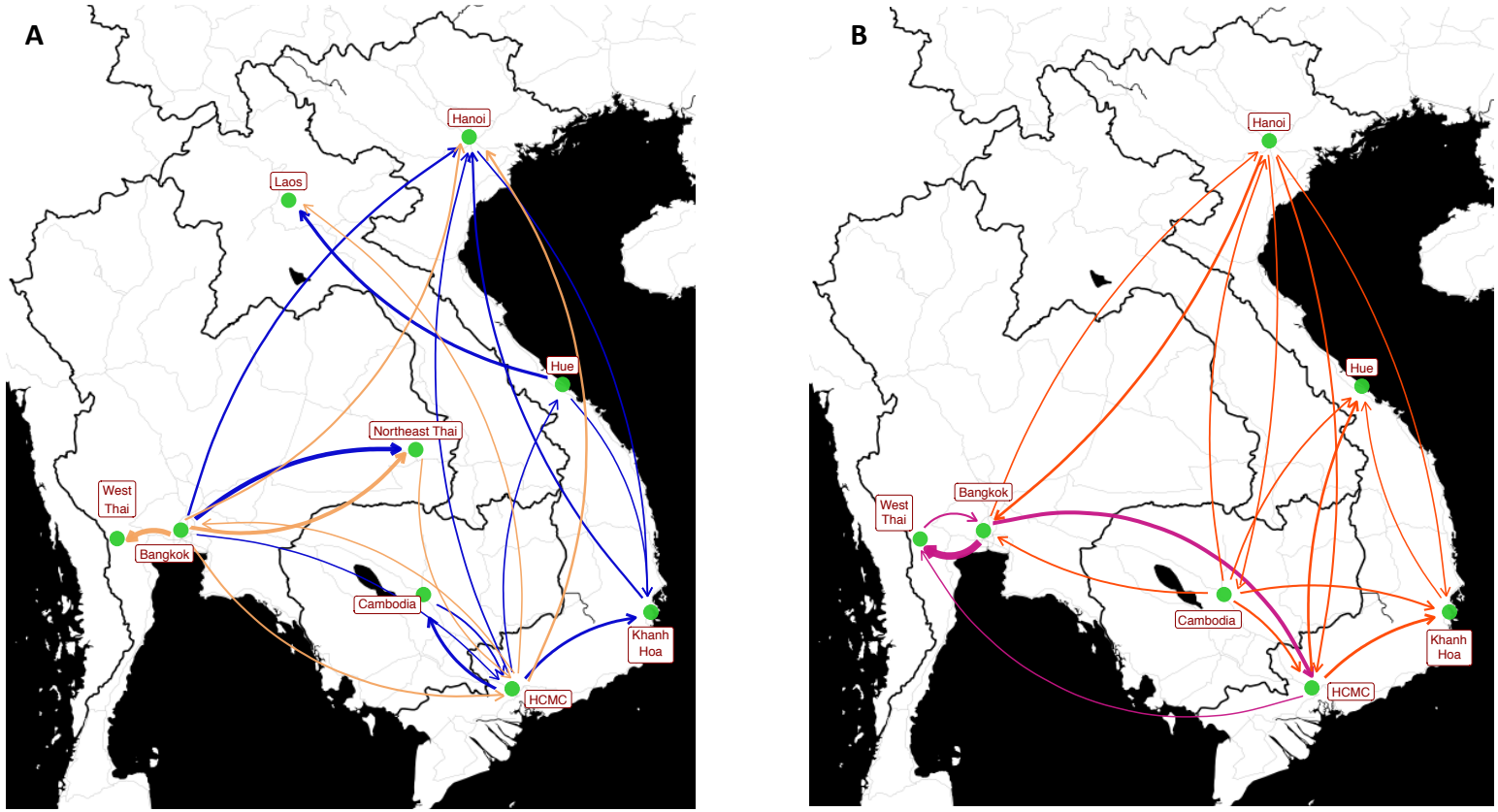


Figure 5.10 Migration network of *Shigella* within Southeast Asia. The figures summarize the geographical transitions of (A) the two *S. flexneri* Lineages, 3.1 (blue arrows) and 3.2 (orange arrows); and (B) the two *S. sonnei* sublineages, Global 3.1 (red arrows) and Global 3.2 (violet arrows). Each region is highlighted by a green circle on the map, with the capitals Vientiane and Phnom Penh selected as representatives for Laos and Cambodia due to the lack of detailed geographical information. The width of the arrows is proportional to the number of transition events between geographical states, which were inferred by stochastic mapping of geographical traits. Transition events less than one are considered insignificant and not represented in this figure.

I aimed to enhance our understanding of the circulation patterns of the various organisms by examining the phylogenetic structures of individual lineages, starting with the ubiquitous *S. flexneri* serotype 2. Genomic analyses revealed that almost all isolates of this serotype fell into the sublineages 3.1 and 3.2. Lineage 3.1 represents a single clonal expansion event unique to SEA. This variant emerged in 1978 and was initially detected in central Thailand through the mid 1980s (Fig 5.11). Though the population declined in significance with time, the lineage continued to propagate via its spread from central Thailand into Vietnam. There, aside from migrations within Vietnam, additional trans-border disseminations further promoted its circulation. This circulation is exemplified by the transfer of *S. flexneri* Lineage 3.1 from HCMC to Cambodia and from Hue to Laos (Fig 5.10 and 5.11). Notably, territorial expansions were accompanied by various serotype conversions, switching from 2a to 2b in one population (in Vietnam and Cambodia) prior to 1993 and from 2a to 1a in another (in Hue and Laos) by 2002 (Fig 5.11). In contrast, SEA *S. flexneri* belonging to Lineage 3.2 were intermingled with organisms isolated in South Asia, Africa, and the Caribbean (Fig 5.11). With its initial diversification dated around the late 1970s, I speculate that Lineage 3.2 represents the global dissemination of *S. flexneri* 2a. However, inadequate chronological and geographical sampling, in terms of global representativeness, did not permit the resolution of the origin of the clade or a complete picture of its geographical distribution. Further, three major independent subpopulations contributed to the sustainment of *S. flexneri* Lineage 3.2 in SEA over the sampling period. One such population was widespread in Central, Western and Northeast Thailand from 1985 to 1995, while the other two lineages exhibited extensive spread in mainland SEA, with a local expansion of *S. flexneri* 2a in Laos around 2002 (Fig 5.11).

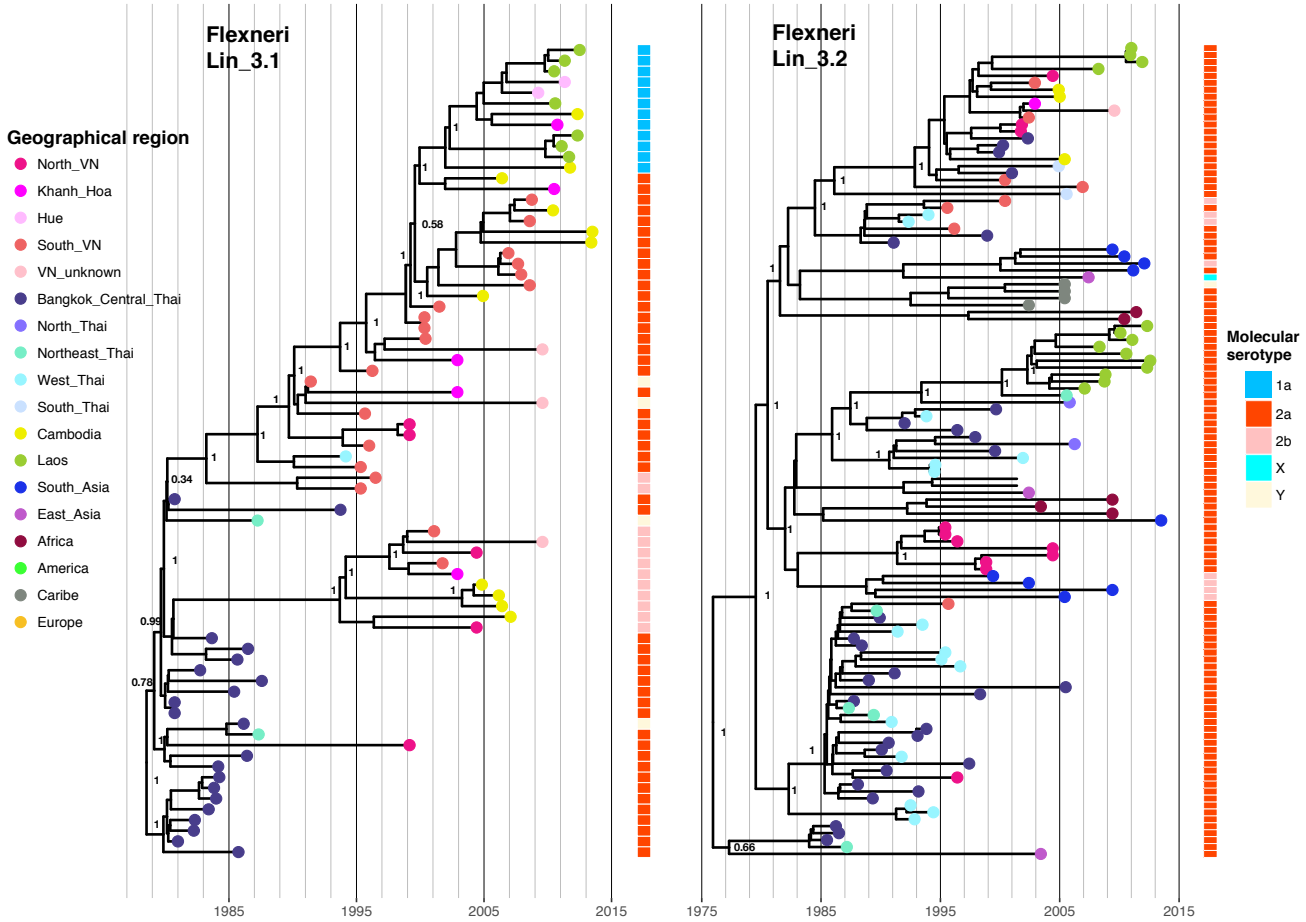


Figure 5.11 Temporal phylogenetic reconstructions of *S. flexneri* Lineage 3. The figure shows the maximum clade credibility (MCC) phylogenies of 76 *S. flexneri* Lineage 3.1 and 117 Lineage 3.2 isolated in Southeast Asia. The leaf colours specify the geographical origins while the column to the right of the tree indicates the molecular serotype of each isolate, as assessed by genomic analyses (see Legend). Posterior probability support is indicated on major internal nodes.

I next investigated the evolution of *S. sonnei* Global 3, which is known for establishing successful and prolonged clonal expansions upon introduction into new populations (Holt *et al.*, 2012). Such expansions in Vietnam have been previously described and named as Vietnam IIIa and IIIb clones, which were demonstrated (by genome-based phylogeny) to be nested within the Global 3.1 and 3.2 clades in this study, respectively. Global 3.1 was also previously studied for its local persistence in South Vietnam (Holt *et al.*, 2013), and I extended its phylogeny by the inclusion of sequences from organisms of neighbouring Laos, Cambodia, and Thailand. This enhanced phylogenetic analysis revealed that *S. sonnei* Global 3.1 was being stably perpetuated in the region as early as in 1983, producing three major subpopulations. Among these, two minor clones were common within Thailand during the late 1990s to the mid 2000s (Fig 5.12). The contemporary Laotian *S. sonnei* mostly clustered into a single group, potentially following an expansion from Thailand between 2000 and 2005. In contrast, trans-border dissemination tended to dominate the transmission pattern of the other major clone, showing that isolates from geographically proximate regions were intermingled rather than forming discrete islands. However, I also observed regional establishment, with the formations of Hue and Khanh Hoa specific circulations from 2008 via the HCMC *S. sonnei* progenitors (Fig 5.10 and 5.12). Additionally, *S. sonnei* was likely imported to SEA on multiple other occasions, as illustrated by the population structure of Global 3.2 (Fig 5.12). Three such introductions were country-specific, often promoting transient clonal amplification in the region. One major introduction in 1986 into Thailand was associated with the expansion of *S. sonnei* into the Central and Western regions. This clone also seeded several exportations into Vietnam.

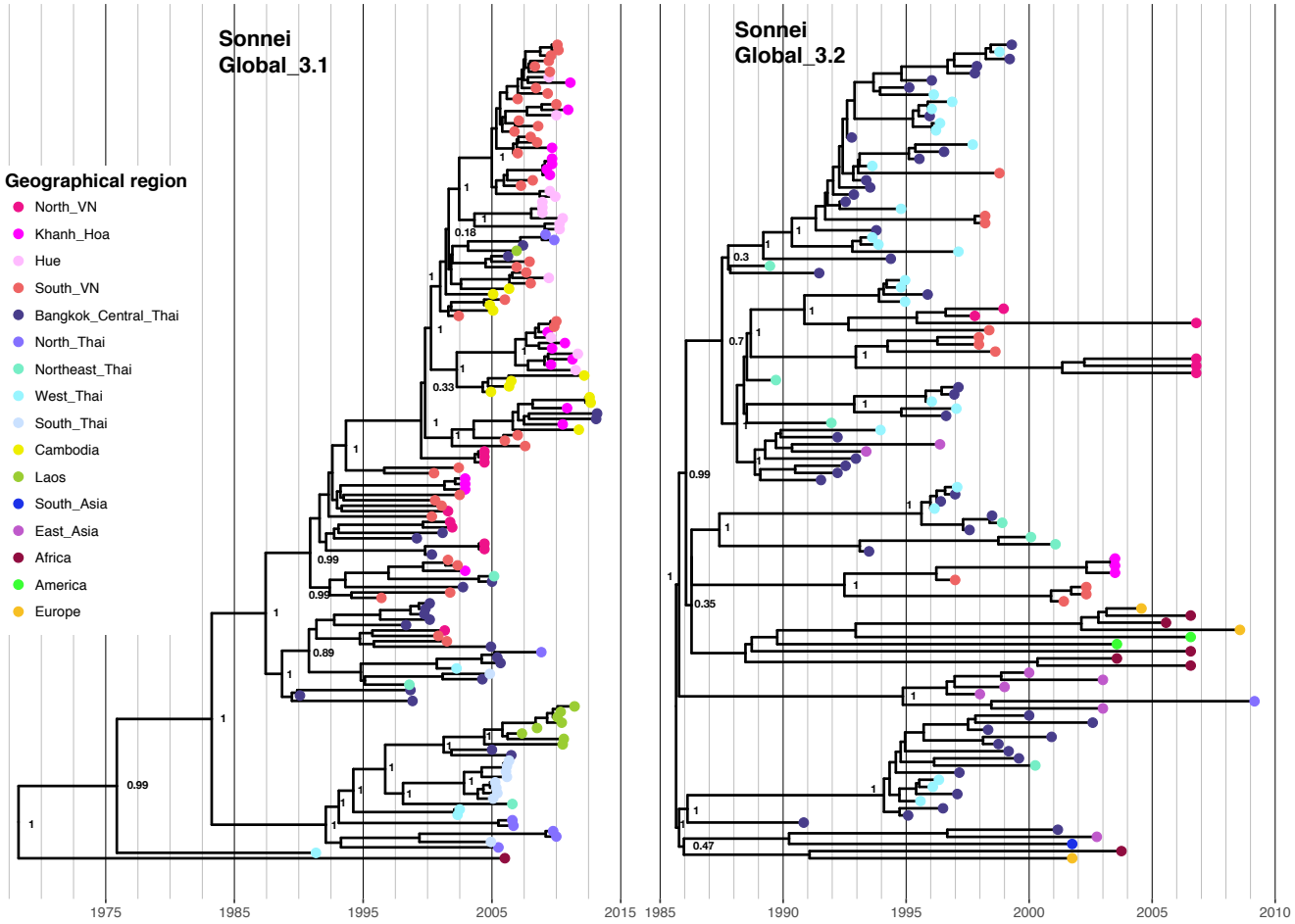


Figure 5.12 Temporal phylogenetic reconstructions of *S. sonnei* Lineage 3. The figure shows the maximum clade credibility (MCC) phylogenies of 151 *S. sonnei* sublineages Global 3.1 and 117 Global 3.2. The tip colours indicate the geographical origins of the isolates (see Legend). Posterior probability support is indicated on major internal nodes.

I additionally evaluated the extent of geographical spread of other *Shigella* lineages by examining their phylogenetic structures. Both SEA *S. flexneri* Lineages 1 and 2 isolates clustered alongside global sequences, mainly with organisms from South Asia, indicating that distinct SEA subpopulations were likely formed recently (1960s-1990s) via introductions from undetermined sources (Fig 5.13 and 5.14). For Lineage 1, these introductions came in several waves, with each undergoing country-specific clonal expansion in Thailand, Vietnam, and Cambodia (Fig 5.13). Serotype 1b is most commonly observed in Lineage 1, but the simultaneous insertion of *gtr*Ic and the loss of *oac*A generated a serotype 1c population around 1975. My data suggest that extant geographically dispersed *S. flexneri* 1c are clonal, and diverged from the ancestral serotype 1b background, rather than from 1a as previously proposed (Tang *et al.*, 2016). The dissemination of this clone into Vietnam in the early 1990s likely explains the surging prevalence of serotype 1c in Northern Vietnam (Stagg, Cam and Verma, 2008). Lineage 2 was composed of organisms almost exclusively of serotype 3a, with limited incidences of serotype conversion (Fig 5.14). Geographical mapping showed that separate populations co-existed in SEA during the sampling timeframe. Especially, trans-border transmission became evident for Lineage 2.2 following likely importation from South Asia to SEA around 1996, resulting in a recent clonal expansion within the region. Where populations were restricted to Thailand (for Lineages 1, 2, 3Xv), isolates from the Central, West and Northeast were intermingled, indicating the prevalence of pan-regional *Shigella* populations in the country.

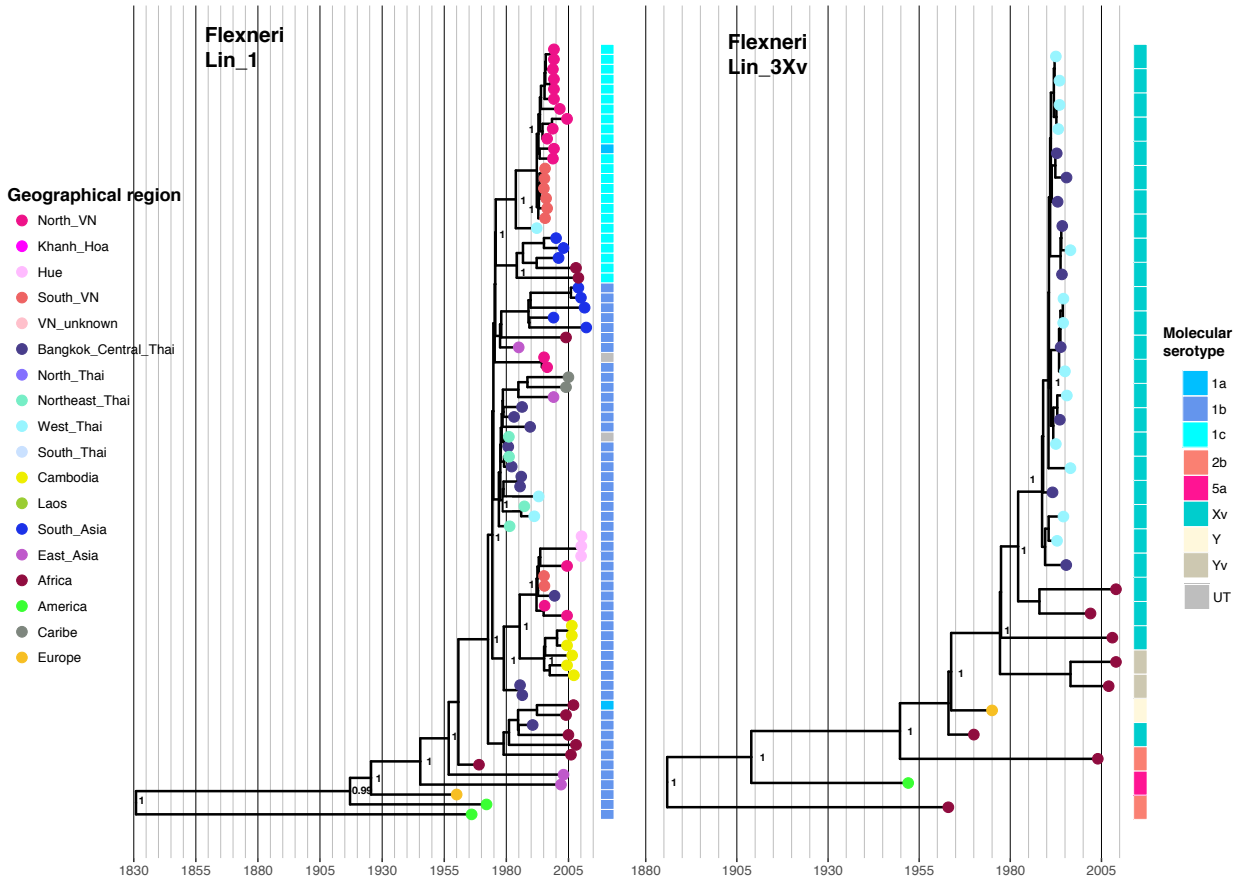


Figure 5.13 Temporal phylogenetic reconstructions of *S. flexneri* Lineage 1 and 3Xv. The figure shows the maximum clade credibility (MCC) phylogenetic reconstructions of 78 *S. flexneri* Lineage 1 and 32 sublineage 3Xv isolated in Southeast Asia. The leaf colours specify the geographical origins while the column to the right of the tree indicates the molecular serotype of each isolate (see Legend). UT: Untypeable. Posterior probability support is indicated on major internal nodes.

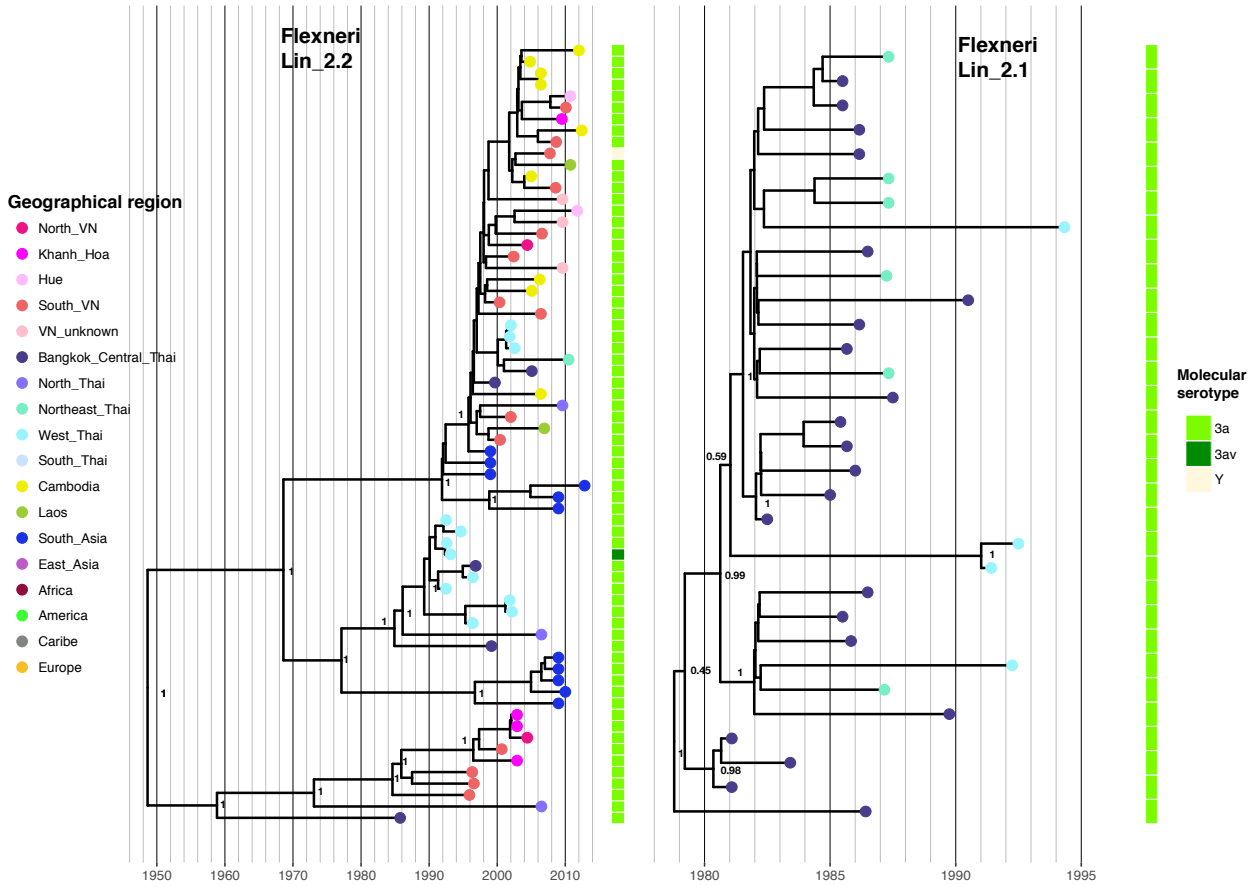


Figure 5.14 Temporal phylogenetic reconstructions of *S. flexneri* Lineage 2. The figure shows the maximum clade credibility (MCC) phylogenies of 32 *S. flexneri* Lineage 2.1 and 68 Lineage 2.2 isolated in Southeast Asia. The leaf colours specify the geographical origins while the column to the right of the tree indicates the molecular serotype of each isolate, as assessed by genomic analyses (see Legend). Posterior probability support is indicated on major internal nodes.

Some in-depth analyses of multiple *Shigella* populations allowed me to compare their evolutionary rates. I found that the nucleotide substitution rate was comparable among lineages and was in the range of 5×10^{-7} to 9×10^{-7} substitutions per site per year (Table 5.1). This range is concordant with previous estimates from global studies of *S. flexneri* (6.46×10^{-7} to 9.54×10^{-7}) and *S. sonnei* (5.2×10^{-7} to 6.7×10^{-7}) (Holt *et al.*, 2012; Connor *et al.*, 2015). However, estimates for *S. flexneri* Lineages 7 and 8 and *S. sonnei* Lineage 1 suggest that these have significantly lower substitution rates, of which 95% HPD do not overlap with those from other *Shigella* lineages (1.4×10^{-7} to 3.84×10^{-7} substitutions per site per year). These lineages also formed the root of the *S. flexneri* and *S. sonnei* global phylogeny respectively and were genetically distant to their more rapidly pervading counterparts (Lineages 1, 2, and 3 for *S. flexneri*, and Global 3 for *S. sonnei*). The estimation may be less accurate for *S. flexneri* Lineage 7, which diverged several hundreds of years ago but only included contemporaneous samples for inference. However, this was not the case for *S. flexneri* Lineage 8 and *S. sonnei* Lineage 1 as they emerged comparatively recently but displayed a disparate evolutionary rate compared to contemporaneous populations. It has been shown in *S. sonnei* that an elevated evolutionary rate is associated with successful global spread and clonal establishment (Holt *et al.*, 2012), as it may facilitate ease of adaptation in new environments. Slower molecular clocks of the aforementioned lineages may be associated with a corresponding reduction in fitness, which may explain their low isolation rate in regional surveillance studies. Specifically, further inspection of *S. sonnei* Lineage 1 revealed that this variant was likely introduced to SEA in the mid-1800s and established transmission locally, which is coupled with a great genetic diversity among isolates. However, its circulation

was restricted mainly to Northern Vietnam, with few isolates from Thailand (Fig 5.15). The most closely related organisms, as well as the immediate sister group, to the SEA Lineage 1 were historical organisms isolated in France. SEA *S. flexneri* Lineage 7, however, shared a distant common ancestor, dating back to the 15th century (Fig 5.16). Nevertheless, circulations within SEA were founded by two major populations that emerged during the mid-1800s, with the dearth of samples detected from outside of the region suggestive of *in situ* evolution and divergence of these two populations within SEA. Genomic interrogation found that 4av was the most common and probably the ancestral serotype for these SEA isolates. Frequent deletions of *gtrIV* and/or pSfYv prompt the conversion to serotypes Yv, 4a, and Y. The more recent *S. flexneri* Lineage 8, emerging in the early 1930s, is commonly associated with serotypes Y and Yv (Fig 5.16). Previous surveillance studies have reported the low prevalence of these serotypes in the region, indicating the minor contributions of Lineages 7 and 8 to the clinical burden of shigellosis (von Seidlein *et al.*, 2006; Bangtrakulnonth *et al.*, 2008; Vinh *et al.*, 2009).

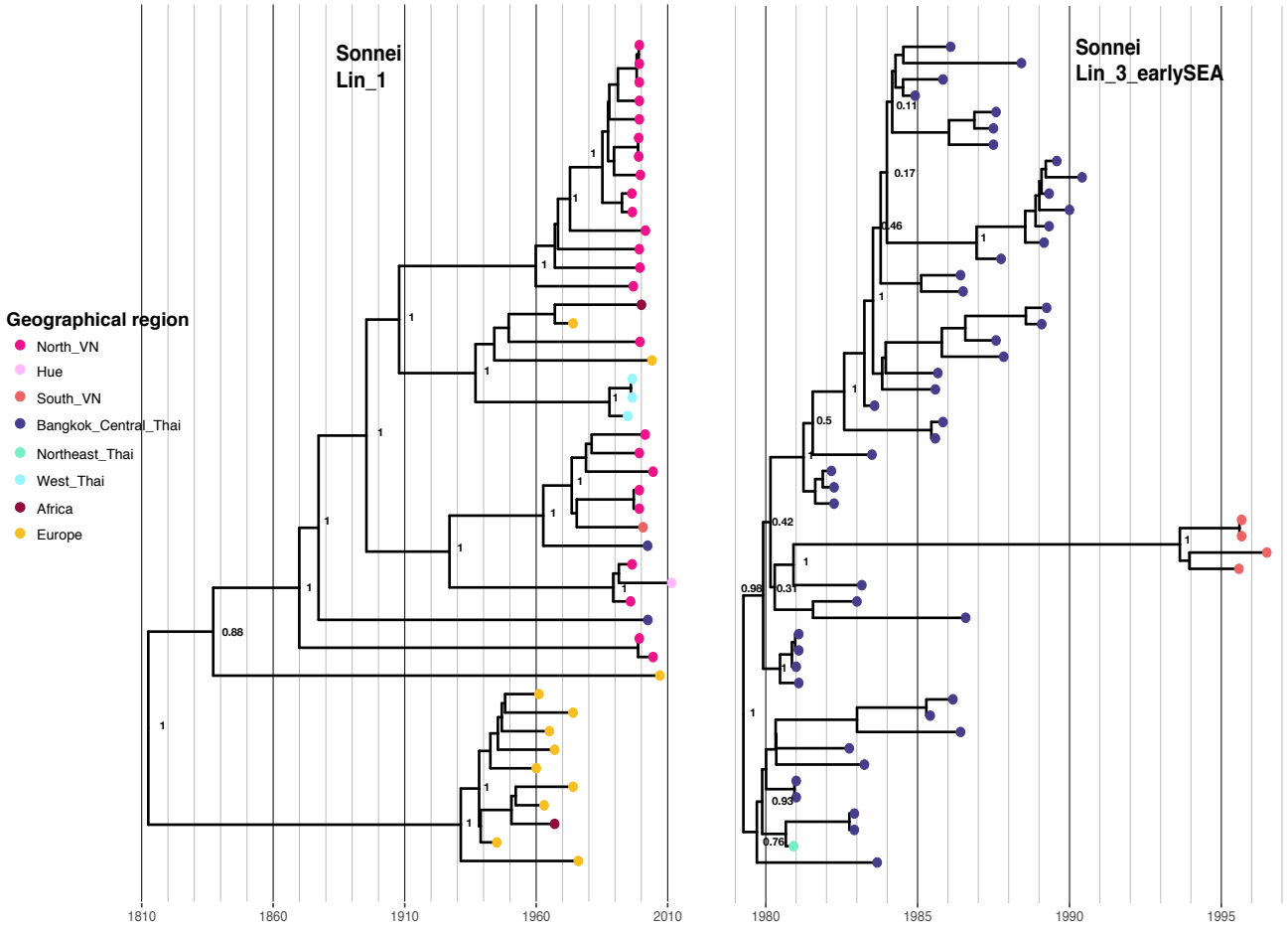


Figure 5.15 Temporal phylogenetic reconstructions of *S. sonnei* Lineage 1 and Lineage 3 clade earlySEA. The figure shows the maximum clade credibility (MCC) phylogenies of 45 *S. sonnei* Lineage 1 and 51 sublineage earlySEA. The tip colours indicate the geographical origins of the isolates (see Legend). Posterior probability support is indicated on major internal nodes.

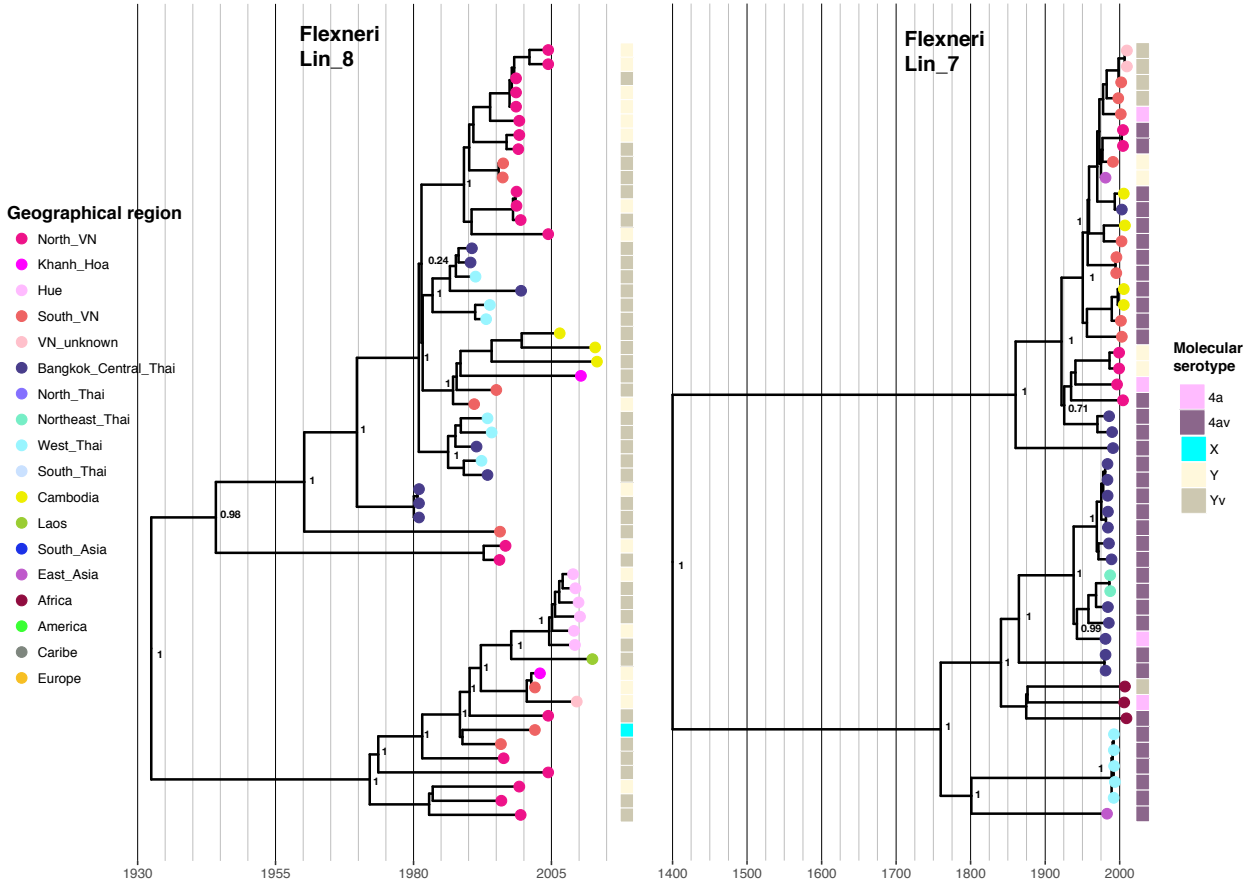


Figure 5.16 Temporal phylogenetic reconstructions of *S. flexneri* Lineages 7 and 8. The figure shows the maximum clade credibility (MCC) phylogenies of 49 *S. flexneri* Lineage 7 and 55 Lineage 8 isolated in Southeast Asia. The leaf colours specify the geographical origins while the column to the right of the tree indicates the molecular serotype of each isolate, as assessed by genomic analyses (see Legend). Posterior probability support is indicated on major internal nodes.

5.3.3 Antimicrobial resistance pattern underlies *Shigella* clonal replacement in Southeast Asia

The diversity and dynamic interplay of different *Shigella* lineages within mainland SEA pose questions regarding the underlying factors governing their ecology. I hypothesised that differences in antimicrobial resistance may play a role in shaping the observed trends. Since antimicrobial susceptibility testing was only available for a subset of isolates, I profiled the AMR genetic determinants (resistomes) of the isolates, supported by the agreement between phenotypic and genotypic results for Enterobacteriaceae. I report the AMR trend over the past three decades for *S. flexneri* and *S. sonnei*, divided by the defined lineages.

Aggregation of nearly one thousand *S. flexneri* isolated across four SEA nations offered unprecedented insight into the regional AMR chronology. Resistance against six antimicrobials was noted in most lineages: streptomycin, ampicillin, chloramphenicol, tetracycline, sulphonamide, and trimethoprim (Fig 5.17), evidencing their maintenance prior to the 1980s. The co-existence of resistance against four of these antimicrobials (streptomycin, ampicillin, chloramphenicol, tetracycline) is associated with the co-organization of their genetic determinants (*strAB*, $bla_{\text{OXA-1}}$, *catA1*, and *tetB*) on the *Shigella* Resistance Locus (SRL) island. This genomic element was independently integrated into multiple *S. flexneri* populations, but not sublineage 3Xv. Thus, this clade showed minimal resistance to sulphonamide, tetracycline, and trimethoprim during its brief period of detection in Thailand. The SRL island was also subject to frequent modification, which was exemplified in the independent deletion of $bla_{\text{OXA-1}}$ and *tetB* in

Lineage 2 since 1990. Notably, there was a marked difference in resistance against trimethoprim pre- and post-1985, with this genotype/phenotype gradually became dominant in all populations (Fig 5.17). This process of resistance acquisition was also evident in the quantity of both *dfrA* variants as they were integrated into this *S. flexneri* collection on a number of occasions. Specifically, the expansion of Lineages 3.2, 3Xv, and part of Lineage 2 were all associated with fixation of *dfrA1*, while the independent acquisition of plasmid-borne *dfrA14* were pervasive in other populations (Lineages 1, 3.1, 7, 8). This distinctive $bla_{\text{TEM-1}}$-*strAB-sul2-dfrA14* resistance integron was found to be co-transferred with a characteristic plasmid backbone. Genomic comparison showed that its structure was almost identical to pRC960-1 (accession number: KY848295), which was recently isolated from *S. flexneri* Y recovered from a porcine stool sample in China (Liang *et al.*, 2018). Independent acquisitions across multiple lineages in all examined countries resulted in the plasmid's presence in nearly 300 *S. flexneri* isolates, but these were generally found more commonly in organisms isolated since the mid-1990s. These results suggest a strong selective pressure in favour of maintaining trimethoprim resistance in *Shigella*. Therefore, I attempted to assess how the genetic dynamics were altered in response to this selective pressure. In order to constrain the pressure exerted by serotype-mediated immune escape, my analysis was independently limited to *S. flexneri* serotypes 2a and 3a isolated in Thailand. For both serotypes, an increase in trimethoprim resistance was obvious and dramatic, particularly leading to fixation in serotype 2a during the early 1990s (Fig 5.18). This trend occurred concurrently with a shift in population structure, with Lin3.1 replaced by Lin3.2 for serotype 2a and Lin2.1 replaced by Lin2.2 for serotype 3a. Alternatively, quinolone resistance in *S. flexneri*, which is

mainly mediated by specific mutations in the QRDR, accelerated significantly after 2000 (Fig 5.17). The selected proliferation of clones bearing specific mutations in the QRDR resulted in predominance of quinolone resistance in more recent isolates. However, these mutations were only apparent in Lineages 2 and 3.1, and to a lesser extent Lineage 8. In contrast, resistance against 3rd generation cephalosporins was only infrequently encountered in Lineages 2 and 3.1 in spite of our vast collection. The patchy appearance noted in the macrolide resistance's heatmap suggests that it emerged rather sporadically over the examined period and demonstrated no clear signal of fixation or selection (Fig 5.17).

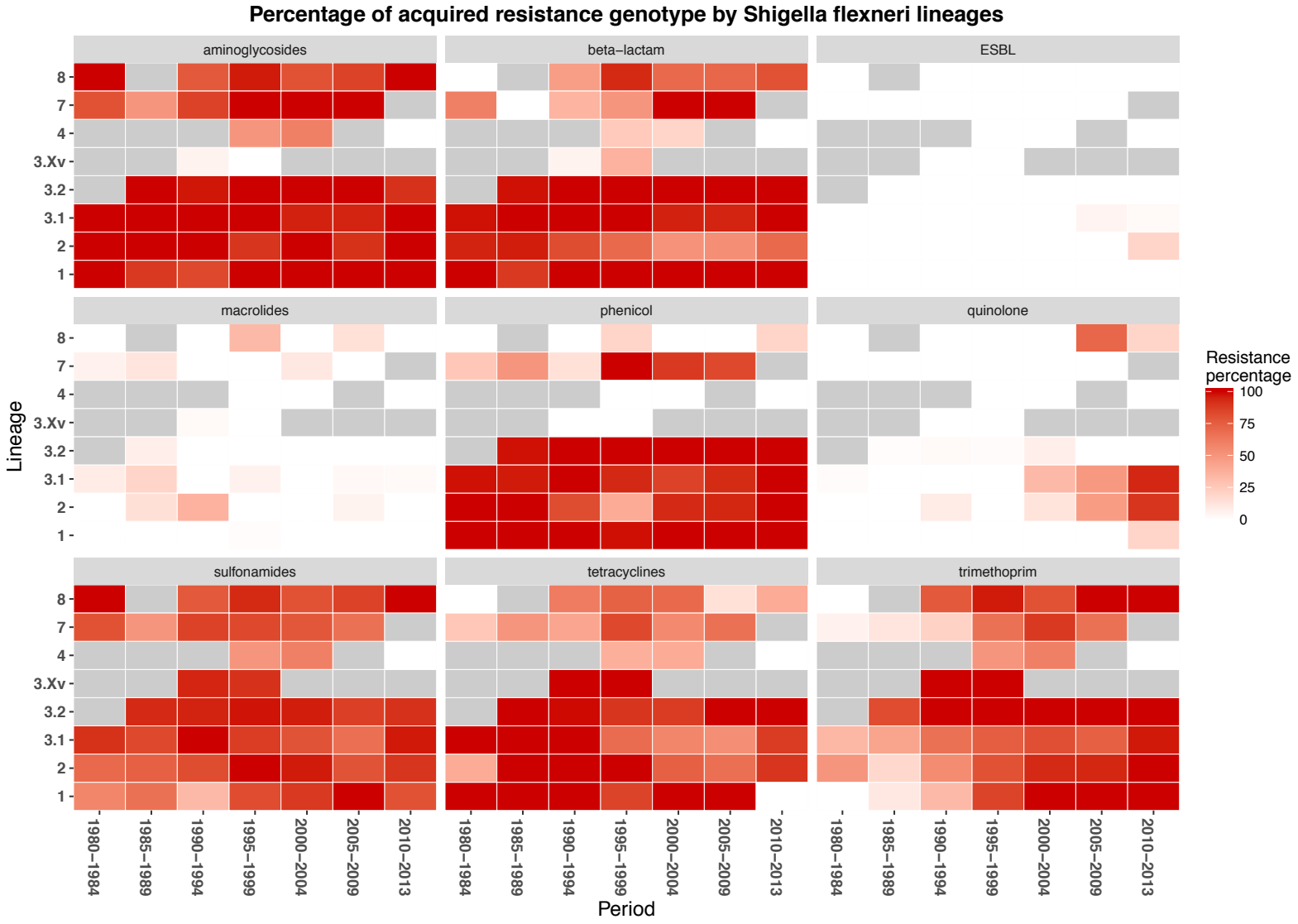
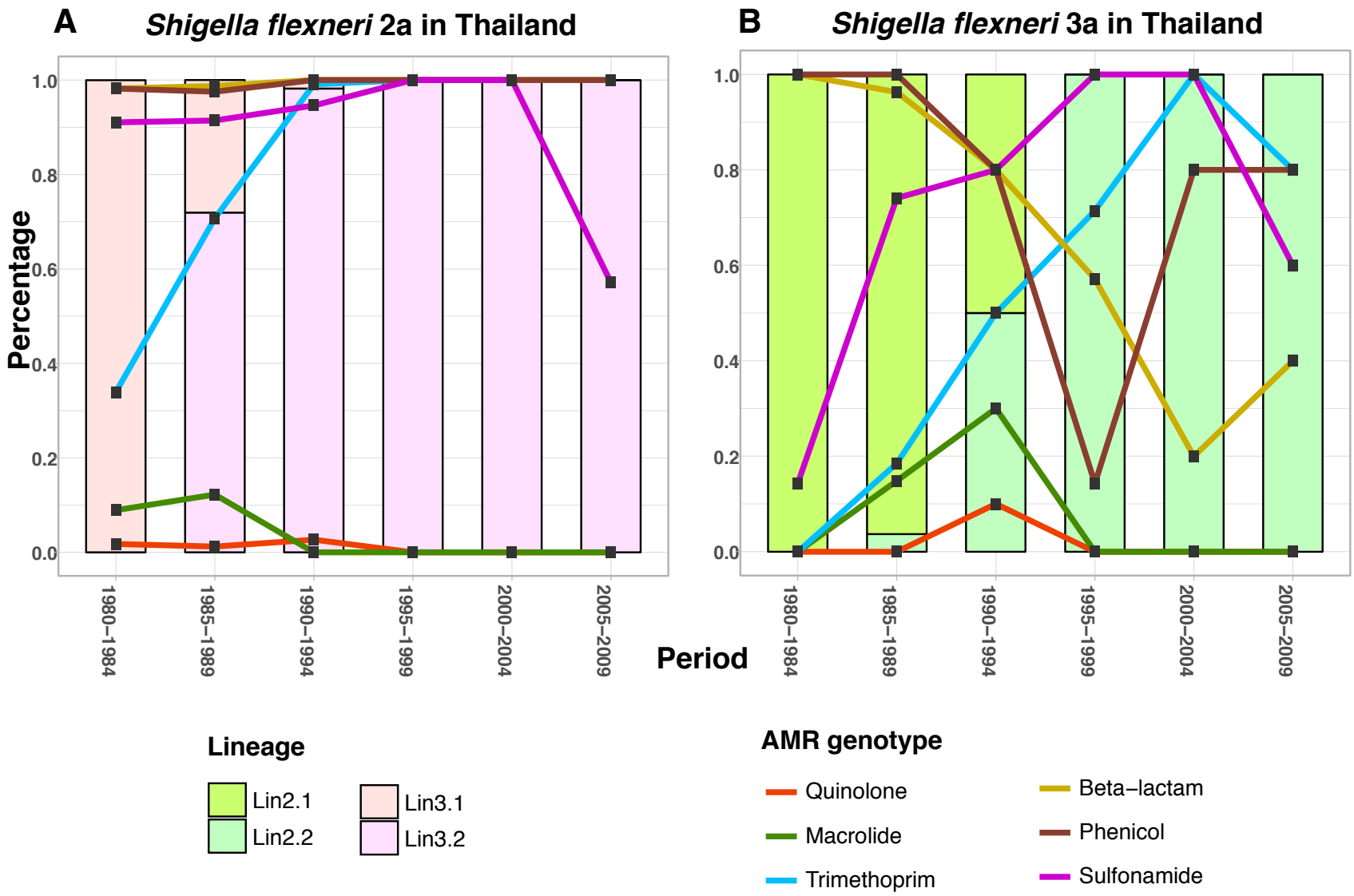

Percentage of acquired resistance genotype by Shigella flexneri lineages
aminoglycosides
beta−lactam
ESBL
macrolides
phenicol
quinolone
sulfonamides
tetracyclines
trimethoprim
Lineage
8
7
4
3.Xv
3.2
3.1
2
1
1980−1984
1985−1989
1990−1994
1995−1999
2000−2004
2005−2009
2010−2013
Period
Resistance percentage
100
75
50
25
0

Figure 5.17 Trend of antimicrobial resistance (AMR) genotype in 990 Southeast Asian *Shigella flexneri*. Each heatmap plot details the resistance trend against a class of antimicrobial, as identified by presence of AMR genetic determinants and categorized into different phylogenetic lineages. The cell's colour intensity positively correlates to the proportion of sequenced isolates belonging to a particular lineage during a defined duration and encoding resistance to the examined antimicrobial class (see Legend). Grey cells denote the absence of sequenced isolates. The antimicrobial classes include aminoglycosides (*strAB*, *aadA*, *aadB*, *aac_3_Ila*), beta-lactam (bla_{TEM}, $bla_{\text{OXA-1}}$), ESBL ($bla_{\text{CTX-M-15}}$, $bla_{\text{CTX-M-24}}$, $bla_{\text{CMY-7}}$), macrolides (*ermB*, *mefB*, *ereA*, *msrE*, *mphA*, *mphE*), phenicol (*cmlA1*, *floR*, *catA1*, *catA2*, *cat3*), quinolones (*qnrS1*, *qepA*, *gyrA* codons S83 or D87 mutations), sulfonamides (*sul1*, *sul2*, *sul3*), tetracyclines (*tetA*, *tetB*, *tetC*, *tetD*, *tetM*), and trimethoprim (*dfrA1*, *dfrA3*, *dfrA5*, *dfrA8*, *dfrA10*, *dfrA12*, *dfrA15*, *dfrA17*).

Figure 5.18 Clonal replacement of *Shigella flexneri* in Southeast Asia. The figures show the isolation frequency of major *S. flexneri* lineages from all regions of Thailand from 1980 to 2009, divided by serotypes with (A) serotype 2a, and (B) serotype 3a. The bar chart displays the proportional sampling of these populations compared to the total *S. flexneri* of the respective serotype isolated during a period (see Legend). The percentage of *S. flexneri* isolates resistant to a given antimicrobial, determined by identification of AMR genetic factors, is plotted by the line graph. The y-axis (percentage) is used to measure both the isolation proportion of *S. flexneri* clades and proportion with resistance to specified antimicrobials.


A
Shigella flexneri 2a in Thailand
B
Shigella flexneri 3a in Thailand
Percentage
1.0
0.8
0.6
0.4
0.2
0.0
1980–1984
1985–1989
1990–1994
1995–1999
2000–2004
2005–2009
Period
Lineage
Lin2.1
Lin2.2
Lin3.1
Lin3.2
AMR genotype
Quinolone
Macrolide
Trimethoprim
Beta–lactam
Phenicol
Sulfonamide

The recent emergence and dominance of *S. sonnei* in developing countries have been coupled with resistance to multiple classes of antimicrobials. Based on a collection of 880 *S. sonnei* sequences from SEA, I examined their resistance genotypes to determine an inter-lineage AMR comparison. I additionally included 71 SEA CenAsiaIII isolates in this analysis, retrieved mostly from Vietnam since 2013, to provide an up-to-date overview. Similar to previous findings, the CenAsiaIII Global 3.1 and 3.2 clades exhibited sustained resistance against trimethoprim, sulphonamide, tetracycline, and streptomycin owing to the inheritance of class 2 integron (*dfrA1*) and the small spA plasmid (*sul2*, *tetA*, *strAB*) during the evolution of Lineage 3 (Fig 5.19). However, this crucial one-time acquisition event took place prior to the divergence of the three aforementioned clades and excluded the early-SEA clade, explaining different AMR profiles of the latter. Specifically, the early-SEA clade stably maintained a genomic island which was almost identical to the *Shigella* resistance locus pathogenicity island (SRL-PAI) described in *Shigella dysenteriae* serotype 1 (Njamkepo *et al.*, 2016). This SRL-PAI encodes resistance to ampicillin ($bla_{\text{OXA-1}}$), aminoglycoside (*aadA*), tetracycline (*tetB*), and chloramphenicol (*catA1*) as well as a *fec* locus for ferric ion uptake. This makes early-SEA's AMR profile more comparable to that of other *S. flexneri* lineages. In addition, *S. sonnei* became resistant to trimethoprim over the sampled period, most notably in the early-SEA clade and Lineage 1. The integration of a class I integron, carrying *aph3-ereA-dfrA5*, into a subset of early-SEA *S. sonnei* during 1989-1990 led to co-resistance against aminoglycoside, macrolide, and trimethoprim. In contrast, Lineage 1 gained another distinctive class I integron in five separate events. This 20 kbp genetic element carried an MDR cassette encoding resistance to trimethoprim (*dfrA12*),

aminoglycoside (*aadA2*), quaternary ammonium (*qacE*Δ), sulphonamide (*sul1*), macrolide (*mphA*), mercury (*merADE*), and ampicillin ($bla_{\text{TEM-1}}$), resulting in a disparate AMR profile of Lineage 1. This element was identical to the one found on an IncFII plasmid pEFC36a isolated from wastewater treatment effluent in Canada (Rahube *et al.*, 2014). In a comparable manner to *S. flexneri*, increases in quinolone resistance in *S. sonnei* were evident in all four clades of Lineage 3 over the years, with this trend most obvious in Global 3.1. This observation was associated with the recent expansion of two clones bearing *gyrA*-S83L (in Khanh Hoa and Hue) and *gyrA*-D87Y (in South Vietnam) mutations.

With the aim of understanding how multiple *S. sonnei* populations co-exist in the region, I tracked their temporal and spatial distributions separately for two countries, Thailand and Vietnam. Since *S. sonnei* Lineage 1 was restricted to Northern Vietnam, it was removed from this analysis. I found that clonal replacement was typical of the dynamics of *S. sonnei* populations in both countries, but the composition of the clonal replacement differed between the two locations, which may be associated with the discrepancy in sampling timespan. In Thailand, the predominance of one particular *S. sonnei* clone lasted for approximately one decade (Fig 5.20A). The early-SEA clade was widespread during the 1980s, but the advent of the 1990s saw a rapid replacement by the Global 3.2. This shift occurred again around the turn of the 21st century, making Global 3.1 the most prominent *S. sonnei* population in Thailand. This circulation dynamic was repeated in Vietnam as the same populations pervaded during similar periods as observed in Thailand. For both examined geographies, clonal alternation brought a dramatic change

in AMR status for *S. sonnei*. For example, resistance to trimethoprim escalated with the arrival of Global 3.2 in Thailand, but phenicol resistance plummeted with the disappearance of early-SEA. Similarly, the increase in resistance against quinolones and 3rd generation cephalosporins was accompanied by the establishment of Global 3.1 in Vietnam (Fig 5.20B). The transient dip in quinolone resistance in the early 2000s signifies that the introduced Global 3.1 was initially susceptible to quinolones but developed resistance mutations during its expansion. The Vietnamese *S. sonnei* population is punctuated by a recent and sudden surge in FQ resistance, which resulted from the local persistence of the introduced CenAsiaIII clade. Resistance to macrolides and 3rd generation cephalosporin also increased recently, indicating that effective sampling from the local AMR gene pool could generate an MDR subclone in a short time period. A substantial increase in resistance against first-line antimicrobials observed in Vietnam contrasts with the population in Thailand, particularly for the quinolones. However, this interpretation needs to be taken with caution because only a handful of Thai isolates were collected after 2006.

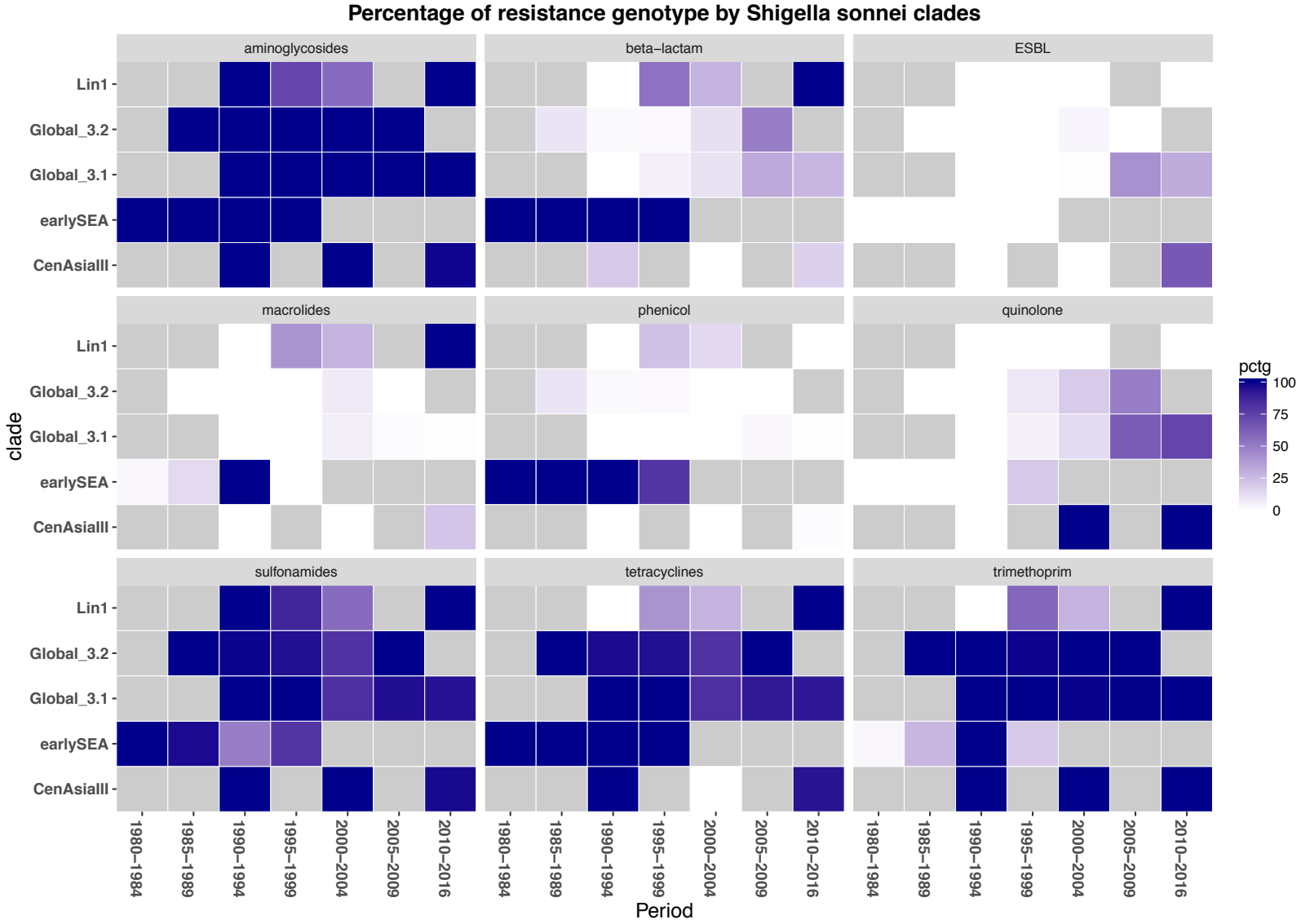
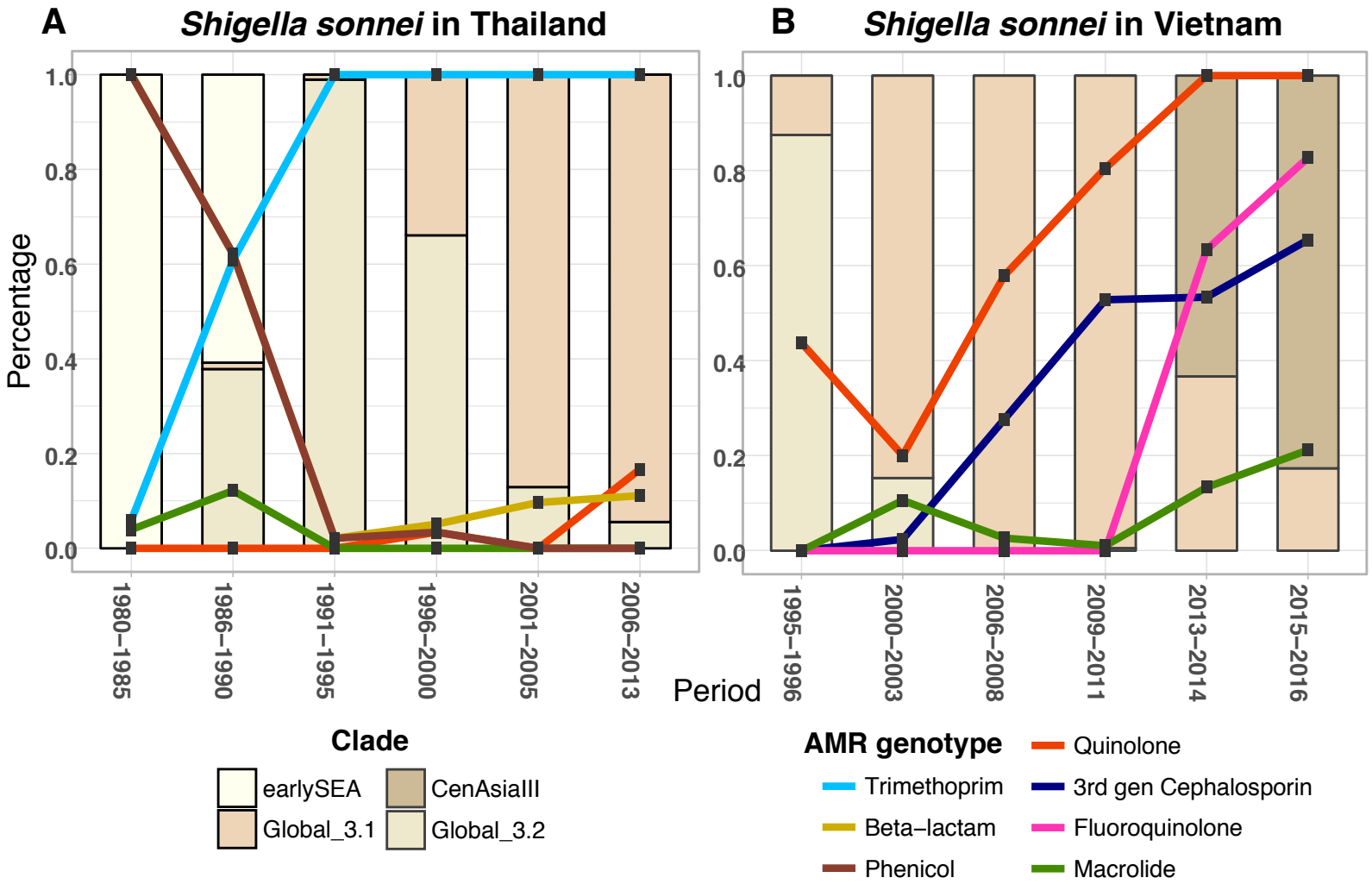

Percentage of resistance genotype by Shigella sonnei clades
aminoglycosides
beta−lactam
ESBL
macrolides
phenicol
quinolone
sulfonamides
tetracyclines
trimethoprim
clade
Lin1
Global_3.2
Global_3.1
earlySEA
CenAsiaIII
1980−1984
1985−1989
1990−1994
1995−1999
2000−2004
2005−2009
2010−2016
Period
pctg
100
75
50
25
0

Figure 5.19 Trend of antimicrobial resistance (AMR) genotype in 880 Southeast Asian *Shigella sonnei*. Each heatmap plot details the resistance trend against a class of antimicrobial, as identified by presence of AMR genetic determinants and categorized into different phylogenetic lineages. The cell's colour intensity positively correlates to the proportion of sequenced isolates belonging to a particular lineage during a defined duration and resistant to the assessed antimicrobial class (see Legend). Grey cells denote the absence of sequenced isolates. The antimicrobial classes include aminoglycosides (*strAB*, *aadA*, *aadB*, *aac_3_Ila*), beta-lactam (bla_{TEM}, $bla_{\text{OXA-1}}$), ESBL ($bla_{\text{CTX-M-14}}$, $bla_{\text{CTX-M-15}}$, $bla_{\text{CTX-M-24}}$, $bla_{\text{CTX-M-55}}$, bla_{DHA}, bla_{CMY}), macrolides (*ermB*, *ereA*, *mphA*), phenicol (*cmlA1*, *floR*, *catA1*), quinolones (*qnrS1*, *qnrB4*, *gyrA* codons S83 or D87 mutations, *parC* S80 mutation), sulfonamides (*sul1*, *sul2*, *sul3*), tetracyclines (*tetA*, *tetB*, *tetC*), and trimethoprim (*dfrA1*, *dfrA5*, *dfrA7*, *dfrA8*, *dfrA12*, *dfrA14*, *dfrA16, dfrA17, dfrA18*).

Figure 5.20 Clonal replacement of *Shigella sonnei* in Southeast Asia. The figures show the isolation frequency of major *S. sonnei* clades in (A) all regions of Thailand from 1980 to 2013 and (B) South and Central Vietnam from 1995 to 2016. The bar chart displays the proportional sampling of these populations compared to the total *S. sonnei* isolated during a period (see Legend). The percentage of *S. sonnei* isolates resistant to a given antimicrobial, determined by identification of AMR genetic factors, is plotted by the line graph. Notice that the timescale on the x-axis differs between the two subplots, and the y-axis (percentage) is used to measure both the isolation proportion of *S. sonnei* clades and resistance proportion to antimicrobials.


A
Shigella sonnei in Thailand
B
Shigella sonnei in Vietnam
Percentage
1.0
0.8
0.6
0.4
0.2
0.0
1980–1985
1986–1990
1991–1995
1996–2000
2001–2005
2006–2013
1995–1996
2000–2003
2006–2008
2009–2011
2013–2014
2015–2016
Period
Clade
earlySEA
CenAsiaIII
Global_3.1
Global_3.2
AMR genotype
Trimethoprim
Beta–lactam
Phenicol
Quinolone
3rd gen Cephalosporin
Fluoroquinolone
Macrolide

5.4 Discussion

The data in this chapter represents the largest and most comprehensive genomic epidemiology study of *Shigella* in SEA ever conducted. With a collection of more than 1,800 sequences from four countries, I was able to document the genetic diversity of various *Shigella* species and serotypes over the past three decades. Heterogeneity in intraspecific populations appeared to delineate both species' circulation patterns, even for *S. sonnei,* which is a single serotype. Within *S. flexneri*, I found that a single serotype is frequently composed of several phylogenetically related but distinct clones, highlighting the difficulty of using serotyping to infer the genetic makeup of the species (Connor *et al.*, 2015). However, this regional dataset revealed that serotype conversion frequency may differ between lineages, with some displaying more extensive serotype-genotype conservation (serotype 3a of Lineage 2) and some experiencing more dynamic transformation (Lineages 3.1, 7, and 8). In some instances, serotype switching was associated with novel regional expansion, exemplified by a change from 1b to 1c in the introduction of Lineage 1 into Vietnam and a change from 2a to 1a in the spread of Lineage 3.1 into Laos and Central Vietnam. I speculate that such events enable better survival for the respective lineage, allowing it to propagate in an immunologically naïve host population and escape competition with other clones of the same parental serotype. In contrast, serotype variations in Lineages 7 and 8 exhibited little evidence of a selective process, and this propensity is instead explained by the stochastic losses of the extra-chromosomal pSFYv.

Despite the diverse genetic makeup of multiple *Shigella* populations, I found that the majority of these progenitors were introduced to the continental SEA during the 1980s-90s. These introductions initially occurred in Thailand's urban areas and disseminated into Vietnam via HCMC and Hanoi, and this pattern was likely driven by increased human movement in the region. It is noticeable that the majority of *Shigella* introductions into Vietnam occurred during the same timeframe, early to mid 1990s. This period witnessed a significant economic transition in Vietnam, which marked its integration into the global market by initiating foreign trade agreements and relaxing trade regulations for the private sector (Thanh, 2005). In addition, heightened domestic migration within Vietnam may have facilitated the human-restricted *Shigella* from South into Central Vietnam, as well as into Laos and Cambodia. Indeed, internal migration increased as economic growth accelerated and migration costs became lower in the 1990s, resulting in the relocation of 4.5 million people (~6% total population) during 1994-99 (Phan and Coxhead, 2010). This figure increased to 8% by the next census in 2009 (Marx and Fleischer, 2010). The trans-border transmission pattern of *Shigella* was most pronounced in populations undergoing extensive clonal evolution in the region, such as *S. flexneri* Lineage 3.1 and *S. sonnei* Global 3.1. Though an intensive sampling strategy was employed to cover both temporal and spatial scales, my interpretations are still subject to the uneven temporal distribution incurred in the collection. Particularly, Thai isolates predominate from 1980-1995 while the Vietnamese isolated were more common from 1995 onward. Organisms from Laos and Cambodia were scarce and only isolated in recent years. These limitations, together with the over-representation of major urban *Shigella* populations in this collection, may bias interpretations. Specifically, the

predominance of Thai isolates in the earlier period might have favoured directional transmissions originated from this country. Estimations of *Shigella* introduction events for each country coincides with its sampling timespan, which might be an artefact of sampling bias. Generally, I observed multiple incoming *Shigella* waves in the 20th century, except for two lineages (*S. sonnei* Lineage 1 and *S. flexneri* Lineage 7) exhibiting earlier regional establishment from the late 19th century. It is surprising that *S. sonnei* Lineage 1 has been maintained for more than a century, which stands in contrast to the transient dominance of other *S. sonnei* populations observed in Southern Vietnam and Thailand. As this lineage is rarely found outside Europe, this distinct epidemiology raises questions related to its exclusive success in Northern Vietnam. It is speculated that the greater genetic diversity recorded within the lineage may also suggest a substantial role of environmental reservoir(s), particularly that facilitates regional waterborne transmission during the last century. While most regions under consideration in these analyses, including Central Thailand, Cambodia, Central and Southern Vietnam, experience a year-round hot weather of the tropics, Northern Vietnam belongs to the humid subtropical climate zone characterized by a clearer seasonal disparity. Meteorological factors have been associated with seasonal fluctuation in shigellosis incidence in endemic countries (Zhao *et al.*, 2016), and I speculate that such climatic differences may affect the survival and transmission of *S. sonnei* in Northern Vietnam. However, future research should examine the validity of this postulation as well as the current epidemiological significance of *S. sonnei* Lineage 1.

Extensive surveillance data have shown that *S. sonnei* is replacing *S. flexneri* as the major aetiological agent of shigellosis in developing regions, including SEA (Bangtrakulnonth *et al.*, 2008; Vinh *et al.*, 2009; Thompson, Duy and Baker, 2015). However, little has been determined regarding the intraspecific population dynamics of these two pathogens. My data show that, for a given locality, *Shigella* epidemiology is characterised by frequent clonal replacement events, and this phenomenon was evident in *S. sonnei* and the major *S. flexneri* serotypes 2a and 3a. Such population alternations coincided with escalating resistance to certain antimicrobials, most notably co-trimoxazole in *S. flexneri* and FQs in *S. sonnei*. This observation led me to speculate that antimicrobial resistance may be a major adaptive strategy in both species. Indeed, the deployment of co-trimoxazole for shigellosis treatment in the 1980s was quickly met with the dominance of its resistance post-1985 in SEA (Niyogi, 2007). Intensified antimicrobial usage coupled with increased human mobility at the global scale may allow *Shigella* to exploit its high degree of genetic variability, at times favouring the settlement of newly introduced but highly resistant clones.

Current evolutionary understanding leads to distinguishing epidemiological behaviour of different *Shigella* species. The epidemiology of *S. sonnei* is driven by intercontinental spread and recent local establishment of pandemic clones (Holt *et al.*, 2012); this trend is apparent in SEA, and exemplified by the local expansion of Global 3.1 and multiple introductions of Global 3.2. In contrast, longer-term local colonization of endemic regions tends to dominate the epidemiology of *S. flexneri* populations, where high genetic diversity is preserved in a location (Connor *et al.*, 2015). This notion is supported by the

occasional recovery of environmental *S. flexneri* and epidemiological associations between this bacterial species and contaminated water (Rahman *et al.*, 2011; He *et al.*, 2012). Moreover, the virulence plasmid of *S. sonnei* is unstable at environmental temperatures, whereas *S. flexneri*'s counterpart is not, signifying that *S. sonnei* may have adapted specifically for person-to-person transmission (McVicker and Tang, 2016). My analysis offers a different perspective on *S. flexneri* evolution in SEA. While the species' population diversity is maintained by the simultaneous existence of multiple *S. flexneri* lineages, these were generally established from the 1950s. This indicates that long-term circulation is uncommon for the majority of *S. flexneri* lineages in SEA. The emergence of new *S. flexneri* clones in the region could be attributed to either: (1) foreign introductions with or without local establishments or (2) successful expansions of particular clones from native *S. flexneri* populations. Recent clonal emergence and the admixture of regional and global isolates in a single lineage lend more support to the former postulation. Thus, I speculate that the introduction-local expansion mechanism has supported the evolution of lineages 1, 2, 3Xv, 3.1 and 3.2. Alternatively, the two ancestral *S. flexneri* lineages, 7 and 8, may represent a scenario where long-term local circulation may play a more prominent role. This collection was limited to clinical samples from the paediatric population, ignoring potential *Shigella* diversity that may arise in environmental reservoirs (for *S. flexneri*) and other shigellosis high risk groups, such as the elderly or the MSM community. To address this limitation, the incorporation of environmental *S. flexneri* should be taken into account in future genomic studies in order to assess their contribution to disease transmission in humans as well as to generate a more comprehensive picture of the evolutionary history of the species.

6 Assessing gut microbiota perturbations during the early phase of infectious diarrhoea in Vietnamese children

All the work detailed in Chapter 6 has been published during my DPhil, under an article in the journal Gut Microbes named:

Chung The, H., Sessions, P. F. de, Jie, S., Thanh, D. P., Thompson, C. N., Minh, C. N. N., Chu, C. W., Tran, T.-A., Thomson, N. R., Thwaites, G. E., Rabaa, M. A., Hibberd, M. and Baker, S. (2017) 'Assessing gut microbiota perturbations during the early phase of infectious diarrhoea in Vietnamese children', *Gut Microbes*. Taylor & Francis, 0(0), pp. 1–17. doi: 10.1080/19490976.2017.1361093.

6.1 Abstract

Diarrhoeal diseases remain the second most common cause of mortality in young children in developing countries. Efforts have been made to explore the impact of diarrhoea on bacterial communities in the human gut, but a thorough understanding has been impeded by inadequate resolution in bacterial identification and the examination of only few aetiological agents. Here, by profiling an extended region of the 16S rRNA gene in the faecal microbiome, I aimed to elucidate the nature of gut microbiome perturbations during the early phase of infectious diarrhoea caused by various aetiological agents in Vietnamese children. Faecal samples from 145 diarrhoeal cases with a confirmed infectious aetiology prior to antimicrobial therapy and 54 control subjects were analysed. I found that the diarrhoeal faecal microbiota could be robustly categorized into four

microbial configurations that either generally resembled or were highly divergent from a healthy state. Factors such as age, nutritional status, breastfeeding, and the aetiology of the infection were significantly associated with these microbial community structures. I observed a consistent elevation of *Fusobacterium mortiferum*, *Escherichia,* and oral microorganisms in all diarrhoeal faecal microbiome configurations, proposing similar mechanistic interactions, even in the absence of global dysbiosis. I additionally found that *Bifidobacterium pseudocatenulatum* was significantly depleted during dysenteric diarrhoea regardless of the aetiological agent, suggesting that further investigations into the use of this species as a dysentery-orientated probiotic therapy are warranted. My findings contribute to the understanding of the complex influence of infectious diarrhoea on gut microbiome and identify new opportunities for therapeutic interventions.

6.2 Introduction

Diarrhoeal diseases result in approximately 1 billion new infections and 500,000 deaths in children aged under five years annually, making it the fourth most common cause of mortality in young children in developing countries (WHO, 2016; Troeger *et al.*, 2017). Repeated diarrhoeal episodes cumulatively increase the risk of malnutrition and stunting, which are associated with cognitive impairment and development of cardiovascular diseases and glucose intolerance in adulthood (Checkley *et al.*, 2008; Deboer *et al.*, 2012; Guerrant *et al.*, 2013). Such syndromes exhaust societal resources, especially in impoverished regions; thus there is a substantial demand for efficacious treatments and prophylaxes.

The human gastrointestinal tract is populated with an immensely rich and diverse microbial communities, with the large intestine harbouring the greatest density of bacteria (The Human Microbiome Project Consortium, 2012). Studies regarding the gut microbiota have highlighted the impact that microbial communities exert on human health, specifically in relation to nutrition (Subramanian *et al.*, 2014), metabolic diseases (Turnbaugh *et al.*, 2006), and cancer (Kostic *et al.*, 2012). Various techniques, including metagenomics, have been employed to investigate microbial disturbances in persistent *Clostridium difficile* infections (Buffie *et al.*, 2014) and inflammatory bowel disease (Morgan *et al.*, 2012), but such dysbiosis remains insufficiently characterized for infectious diarrhoea in high incidence settings. The highly dynamic succession of microbial colonization in young children confounds analysis in this target group (Bäckhed *et al.*, 2015; Dogra *et al.*, 2015). Nevertheless, several 16S rRNA gene profiling studies have shown consistent patterns in the initial gut microbiota response following acute diarrhoeal episodes, in which a transition toward Proteobacteria and *Streptococcus* is observed in an increasingly oxygenated environment (Monira *et al.*, 2013; Pop *et al.*, 2014; David *et al.*, 2015). This shift is additionally coupled with a reduction in specific Firmicutes and Bacteroidetes colonizers; their relative abundance is restored during post-diarrhoeal recovery state (Monira *et al.*, 2013; Hsiao *et al.*, 2014; Becker-dreps *et al.*, 2015; David *et al.*, 2015). These previous studies have either focused on diarrhoea associated with *Vibrio cholerae* (Monira *et al.*, 2013; Hsiao *et al.*, 2014), or did not offer sufficiently detailed granularity for understanding microbiome alternation (Pop *et al.*, 2014; Becker-dreps *et al.*, 2015).

Here, I aimed to characterize modifications of the gut microbiota in the early phase of infectious diarrhoea in Vietnamese children by examining their faecal bacterial 16S rRNA composition. The samples included in this study were a subset originating from a previously described prospective observational study of paediatric diarrhoea carried out in HCMC, Vietnam (Thompson *et al.*, 2015).

6.3 Results

6.3.1 16S rRNA gene sequencing of faecal diarrhoeal samples

I subjected a collection of 200 faecal samples (55 samples from non-diarrhoeal controls and 145 samples from diarrhoeal patients) from Vietnamese children, to DNA extraction and 16S rRNA gene sequencing (Table 6.1). 15/54 controls showed presence of at least one pathogen after detection, including Norovirus, Rotavirus, *Salmonella*, and *Campylobacter*. However, the isolation of these organisms in asymptomatic carriers is frequently observed in endemic setting (Thompson *et al.*, 2015). One hundred and ninety nine of the faecal samples produced 16S rRNA sequences (median library size =2,880,968 paired-end reads). Subsampling and quality filtering generated sub-libraries with a median depth of 756,829 paired-end reads, which served as the inputs for EMIRGE assemblies of the V3-V6 region. A mean of 74% of the reads (range: 37% to 92.5%) were successfully mapped and retrieved. A total of 131,702 assembled and clustered sequences, representing a pool of individual sample OTUs, were produced from all samples. Subsequent filtering, OTU reconstruction, and chimeric sequence removal resulted in 7,479 OTUs clustering at 97% similarity in the 199 faecal samples.

Table 6.1. Summary of the aetiologies associated with 142 diarrhoeal samples contributing to this study

Aetiological agent	**Number**	**Proportion (%)**
Viral	39	27.5
Norovirus (NoV)	19	13.4
Rotavirus (RoV)	20	14.1
Bacterial	80	56.3
Campylobacter	14	9.9
Salmonella	15	10.6
Shigella	49	34.5
Plesiomonas	2	1.4
Mixed	23	16.2
Campylobacter + NoV/RoV	7	4.9
Plesiomonas + NoV/RoV	1	0.7
Salmonella + NoV/RoV	4	2.8
Shigella + NoV/RoV	11	7.7
Total	142	100

6.3.2 Microbial structures within the faecal samples

I examined the microbial composition of the faecal specimens from controls and diarrhoeal cases using the relative abundance of 205 agglomerated unique genera and examined their phylogenetic relationships. *De novo* clustering of the weighted Unifrac dissimilarity matrix, representing the pairwise diversity between samples, segregated the samples into four community state types (CSTs). The optimal number of CST groups was determined by gap statistics (Fig 6.1) and supported by a prediction strength of >75%. A principal coordinate analysis (PCoA) using the weighted Unifrac matrix provided further support for the structuring of the four different microbial CSTs, with the first three coordinates accounting for approximately 70% of total explained variance (Fig 6.2). Constrained analysis on the principal coordinates of 195 samples with their associated metadata indicated that age, weight-for-age Z (WAZ) score, and disease status (diarrhoea/asymptomatic) best explained the variation between CSTs (Fig 6.3). A random forest algorithm was employed to reclassify each sample to one of the four CSTs based on the relative abundance of 205 genera, yielding a predicted error rate of 11.06%. Twenty-eight genera with the highest calculated variable importance measurement (>0.002) were selected as the most significant taxa segregating the four CSTs. Random forest classification was then repeated with these taxa only to re-assess the most predictive genera for CST clustering, yielding a similar misclassification error rate of 11.56% (Fig 6.4 and Fig 6.5). The following taxa predominated in each of the CSTs (with species included based on their abundances): CST1, *Bifidobacterium*-dominant (*B. longum*, *B. breve*, *B. pseudocatenulatum*, *B. bifidum*); CST2, *Bacteroides*-dominant (*B. fragilis*, *B. vulgatus*, *B. uniformis*, *B. caccae*); CST3, *Streptococcus*-dominant (*S.*

gallolyticus and *S. salivarius*); and CST4, *Escherichia*-dominant (Fig 6.6). A multivariate homogeneity of group dispersions test showed that CST3 possessed the highest measurement. This high variability was explained by the fact that a number of samples from this CST were rich in *Megamonas*, as opposed to *Streptococcus*. For diarrhoeal cases, the median reported duration of diarrhoea onset was two days (IQR: 1-3 days). Thus, I classified my findings as representing the early phase of infectious diarrhoea since the gut microbiota composition differs as diarrhoea progresses to later stages (Hsiao *et al.*, 2014; David *et al.*, 2015).

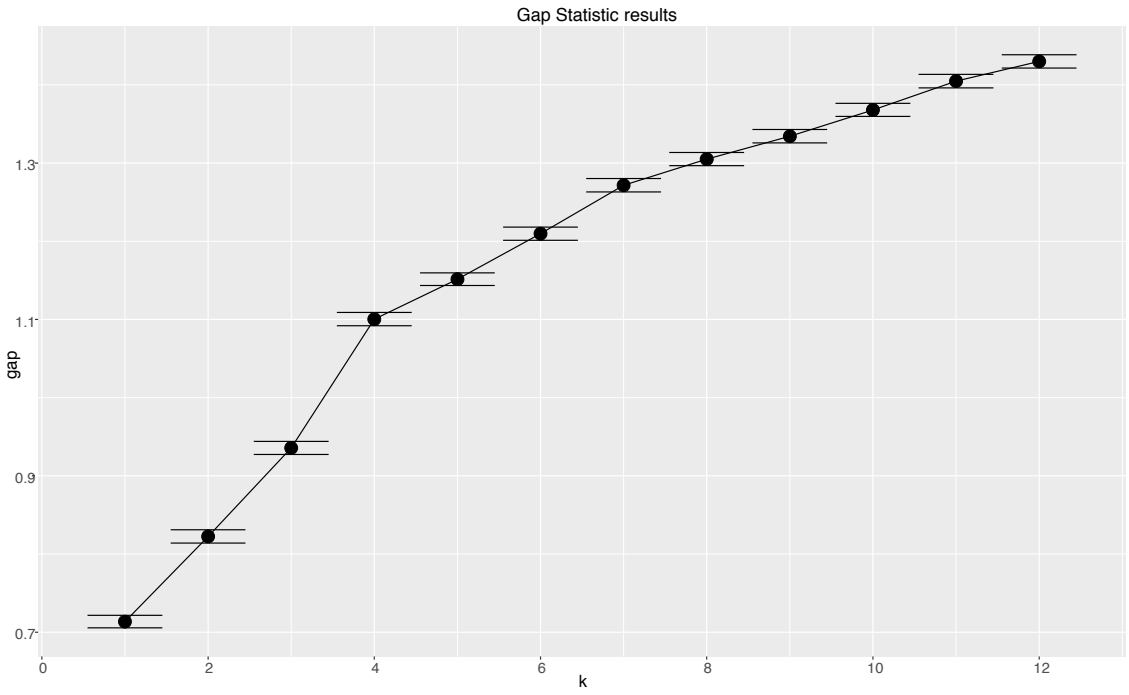


Figure 6.1 Estimating the number of clusters for the microbial compositions of 199 faecal samples

Partitioning around meloids (pam) clustering was performed on the genus-level agglomerated weighted Unifrac dissimilarity matrix of 199 samples. The gap statistic, a goodness of clustering measure, is calculated with different cluster number (k; k-max=12) using the pam clustering outputs. Four was selected as the optimal number for clustering the microbial composition from 199 stool samples, as indicated by a noticeable change in the slope.

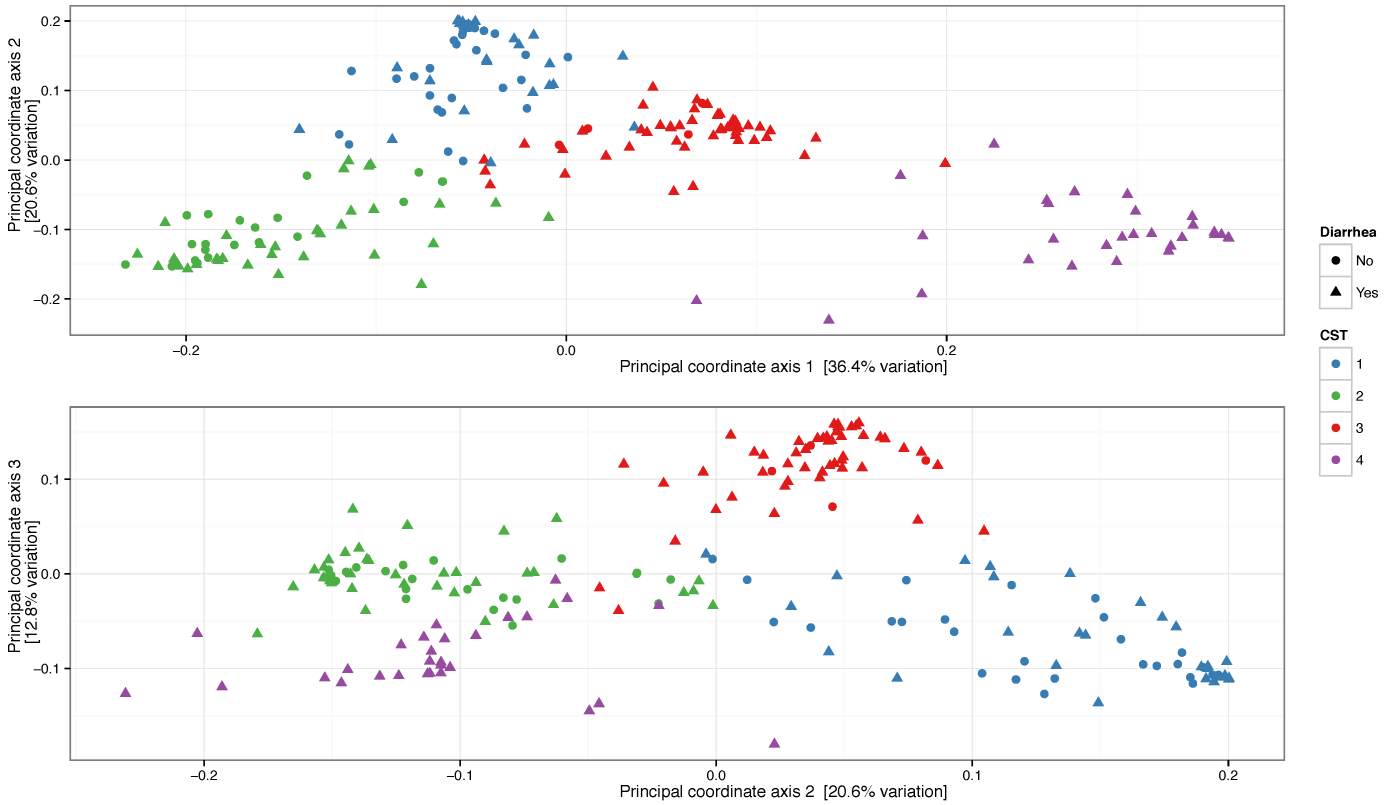


Figure 6.2 Dissimilarity in bacterial compositions of four Community State Types (CSTs)

Principal coordinate analysis (PCoA) was performed on the genus-level agglomerated weighted Unifrac dissimilarity matrix of 199 faecal samples. The similarities in gut microbial composition from non-diarrhoeal controls and diarrhoeal patients in four CSTs (see legend) were projected on the first three principal coordinate axes.

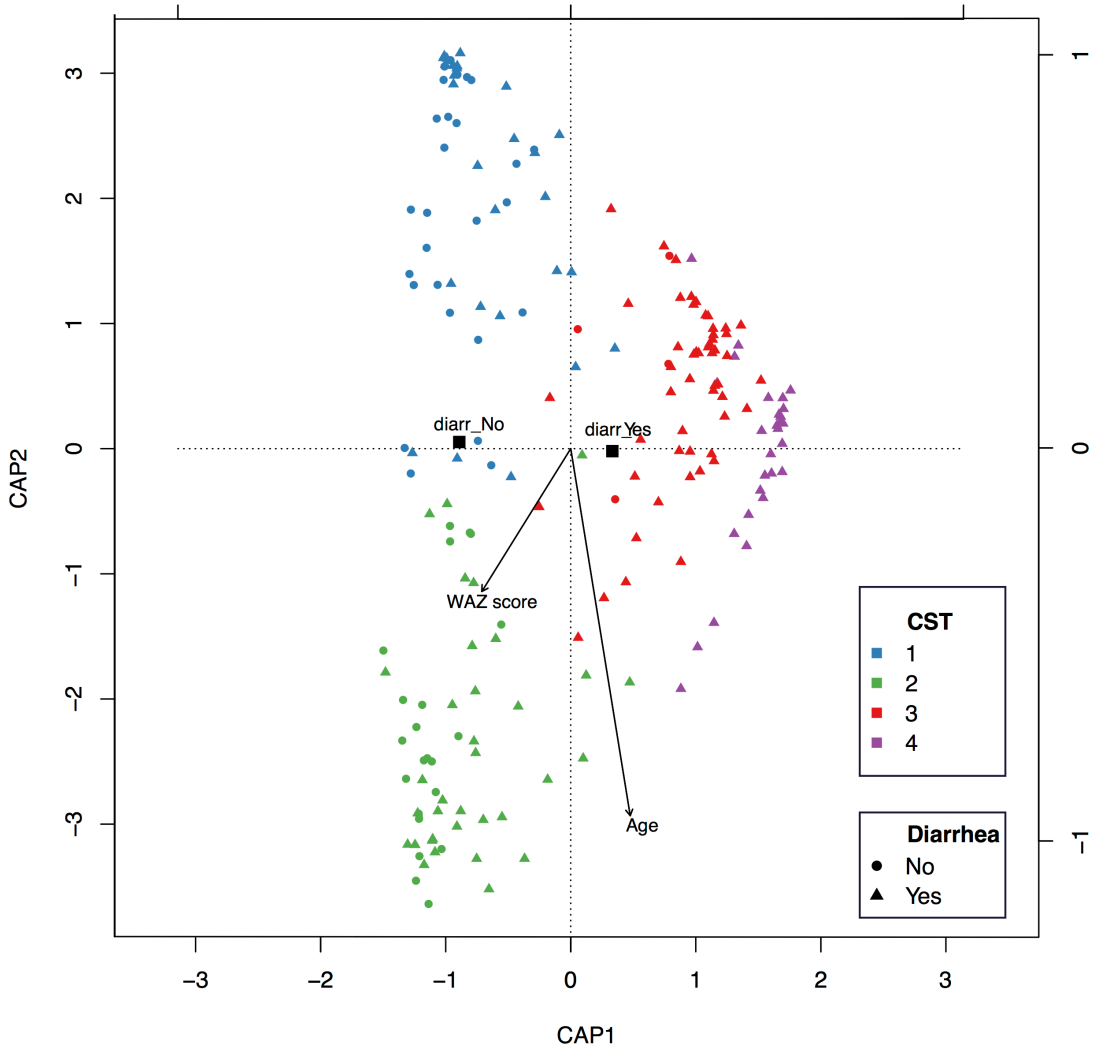


Figure 6.3 The impact of diarrhoea and demographics on the gut bacterial configurations. Constrained analysis of principal coordinates (CAP) biplot displaying the relationship between bacterial compositions (coloured circles and triangles; see legend) and selected metadata. CAP was performed on the genus-level agglomerated weighted Unifrac dissimilarity matrix of 195 faecal samples with complete metadata. Black arrows denote the magnitudes and directions of quantitative demographic variables (age and WAZ score), and black squares represent qualitative variables (centred around relative diarrhoea status).

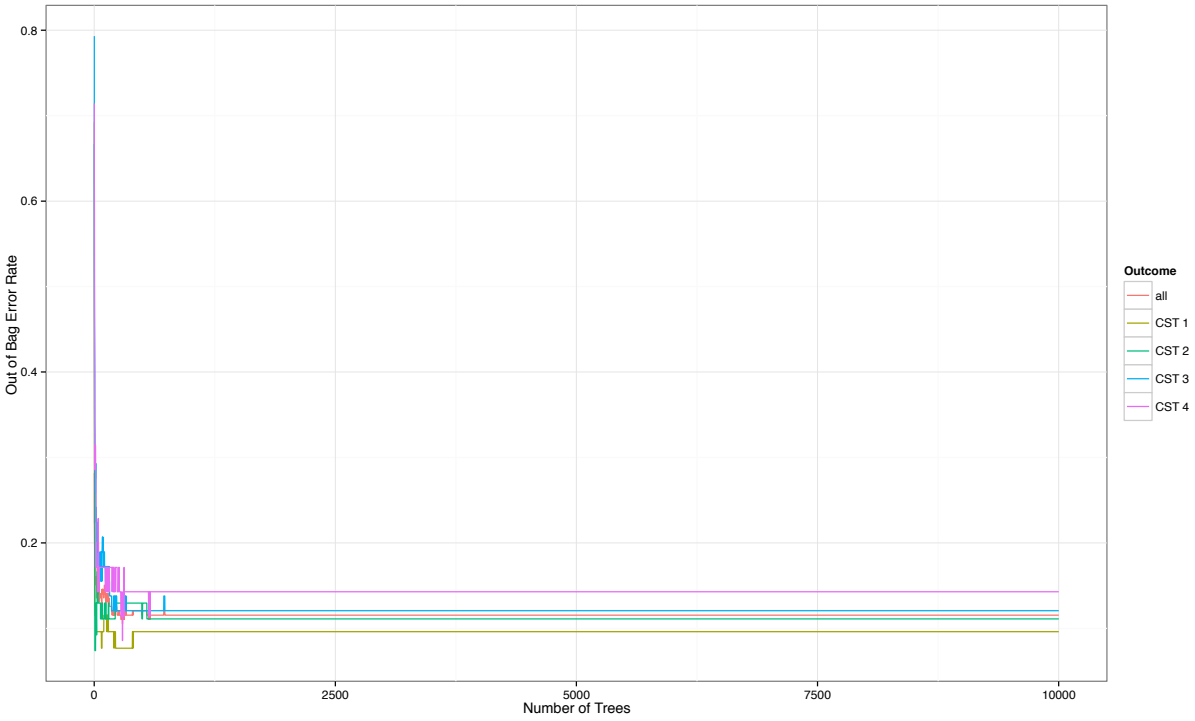


Figure 6.4 Examination of CST clustering with Random Forest classification

Random Forest classification was performed on the relative abundances of 28 bacterial genera with highest calculated variable importance against the pre-defined CST memberships of 199 samples. Out of bag (OOB) error rates of the classification performance against the number of trees used, calculated for all samples and each CST (see legend).

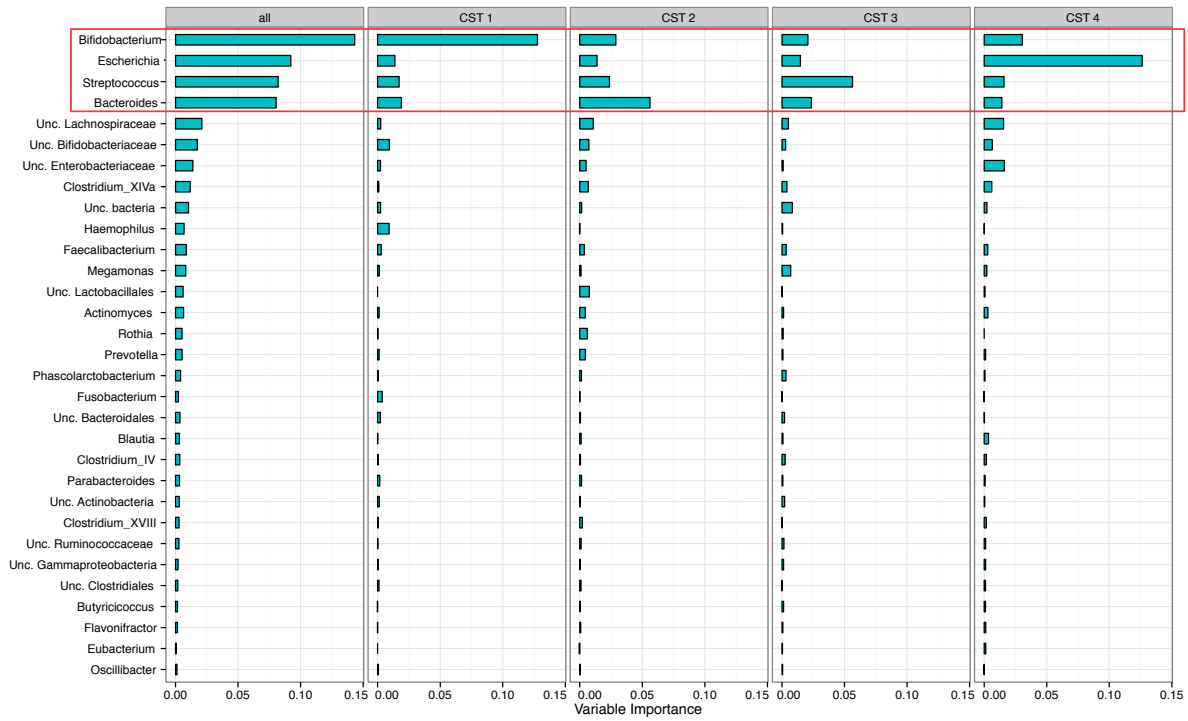


Figure 6.5 Genera defining the clustering of CSTs

Twenty-eight CST-discriminatory bacterial genera identified with highest calculated importance in the Random Forest classification model, shown for all CSTs and within each CST. Genera within the red rectangular box are taxa delineating the clustering of bacterial composition into four CSTs.

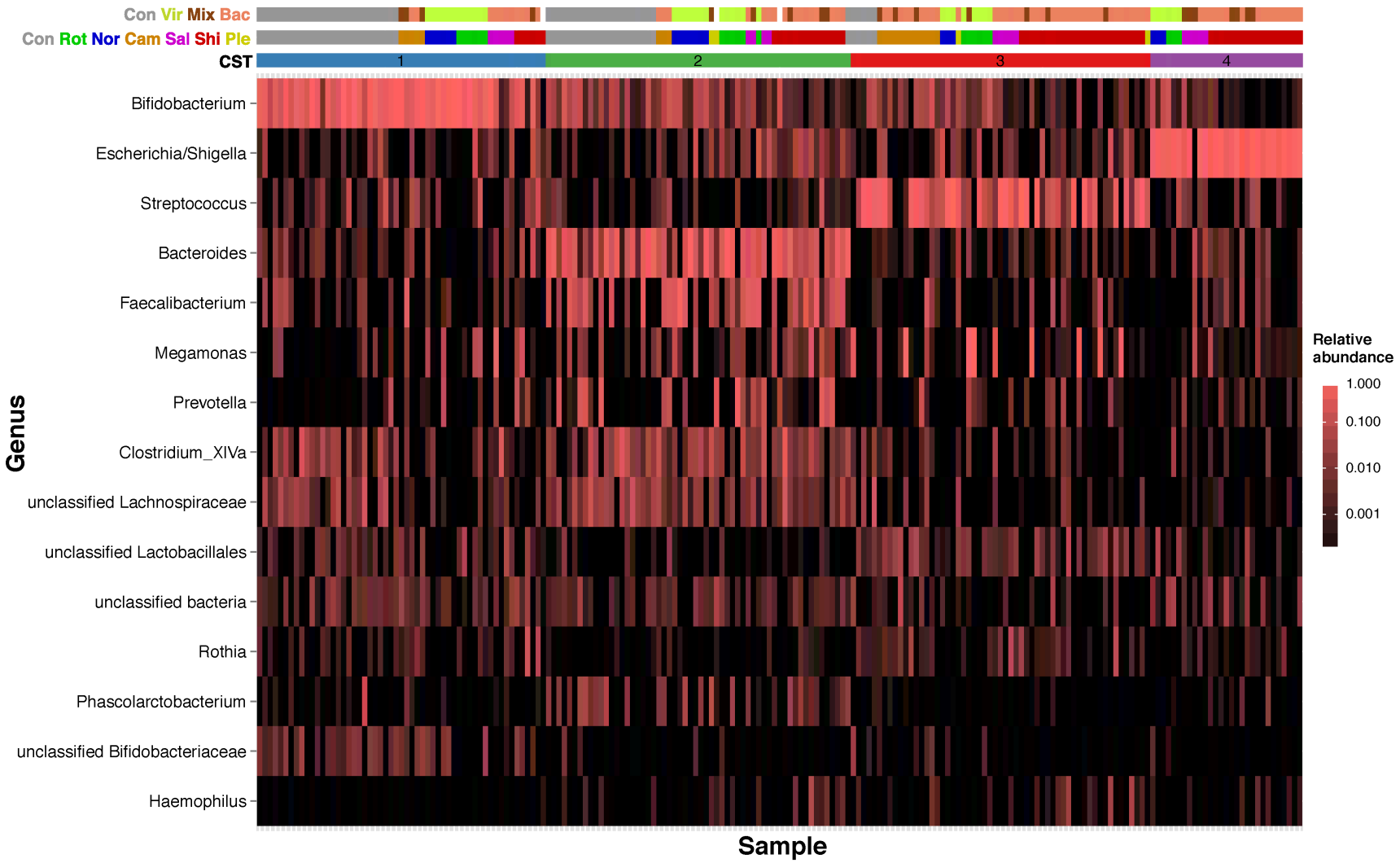

Con Vir Mix Bac
Con Rot Nor Cam Sal Shi Ple
CST
1
2
3
4
Bifidobacterium
Escherichia/Shigella
Streptococcus
Bacteroides
Faecalibacterium
Megamonas
Prevotella
Clostridium_XlVa
unclassified Lachnospiraceae
unclassified Lactobacillales
unclassified bacteria
Rothia
Phascolarctobacterium
unclassified Bifidobacteriaceae
Haemophilus
Genus
Sample
Relative abundance
1.000
0.100
0.010
0.001

Figure 6.6 The bacterial compositions of the four community state types (CSTs) Heatmap showing the fractional abundance of the 15 most important genera determining the clustering of 199 diarrhoeal and non-diarrhoeal children faecal samples, as defined by random forest classification. Clustering was performed on the genus level weighted Unifrac dissimilarity matrix (with 205 genera) using the partitioning around meloids algorithm and identified four CSTs. The nature of sample is indicated at the head of the diagram: non-diarrhoeal control (grey); viral infection (chartreuse); mixed bacterial and viral infection (brown); bacterial infection (salmon); missing data (white). The confirmed major aetiologies of infection are indicated in the middle row: non-diarrhoeal control (grey); Rotavirus (green); Norovirus (blue); *Campylobacter* (orange); *Salmonella* (magenta); *Shigella* (red); *Plesiomonas* (yellow).

The majority of the asymptomatic samples were within CST1 (n=27) and CST2 (n=22), which were generally enriched for taxa known to correspond with a healthy gut microbiota in young children (Koenig *et al.*, 2011; Bäckhed *et al.*, 2015). CST2 harboured a greater degree of bacterial diversity than CST1 and encompassed various genera that delineate an increasingly mature gut microbiota, including *Faecalibacterium*, *Prevotella*, *Clostridium*, Lachnospiraceae, and *Phascolarctobacterium* (Subramanian *et al.*, 2014; Bäckhed *et al.*, 2015). The faecal samples from the diarrhoeal cases were distributed across the four CSTs, with CST3 (n=52), and CST4 (n=29) comprised almost exclusively of diarrhoeal cases. This structure corresponded with the dominance of *Streptococcus* and *Escherichia* species in gut microbiota of diarrhoeal children, as reported previously (Pop *et al.*, 2014; David *et al.*, 2015). The estimated Shannon diversity index was elevated in both controls and the cases in CST2 in comparison to the other CSTs ($p<0.05$, ANOVA-Tukey's test) (Fig 6.7). I additionally observed that bacterial diversity was lower in the diarrhoeal cases relative to the controls within CST1 and CST2, however this difference between the groups was not statistically significant ($p>0.05$, ANOVA-Tukey's test).

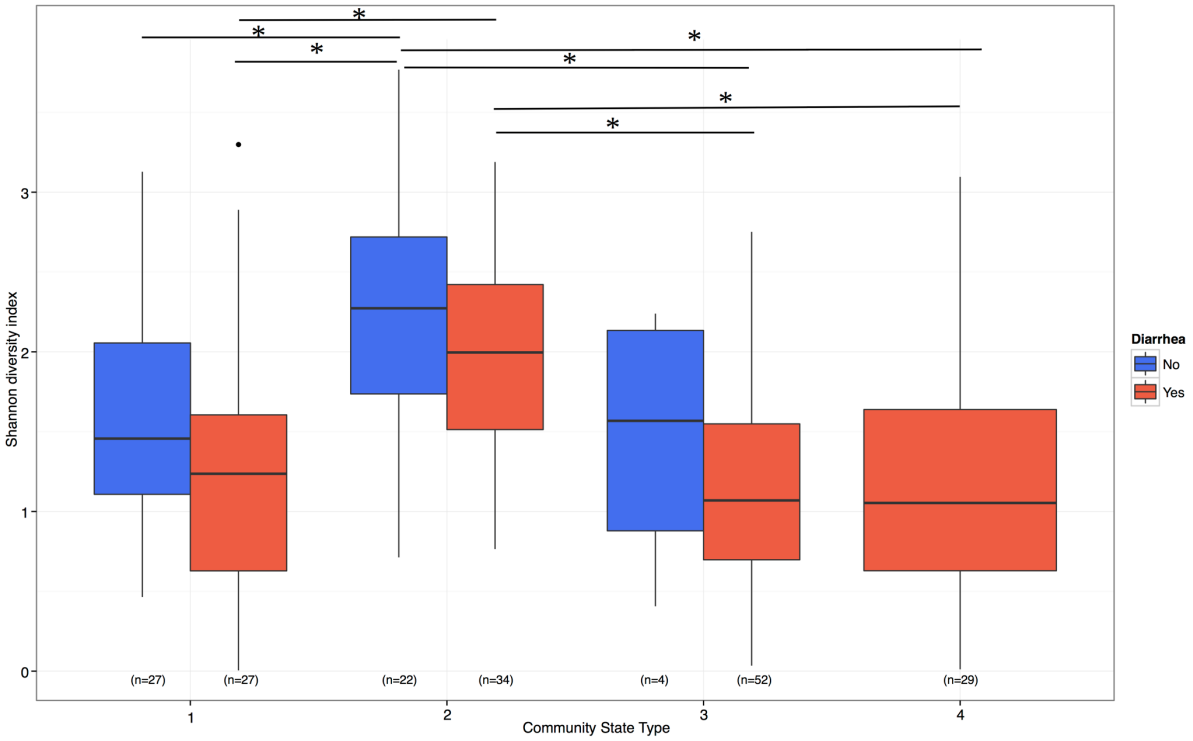


Figure 6.7 Shannon diversity index for 195 faecal samples, classified by diarrhoeal status and CST membership

Boxplots showing the Shannon diversity index, the upper whisker extends from the 75th percentile to the highest value within the 1.5 * interquartile range (IQR) of the hinge, the lower whisker extends from the 25th percentile to the lowest value within 1.5 * IQR of the hinge. Data points beyond the end of the whiskers are outliers. The asterisk indicates statistical significant in the pairwise comparison between CST2 and the other CSTs (ANOVA-Tukey's test, p<0.05).

6.3.3 Factors affecting microbial structure in the faeces of children with diarrhoea

The CST grouping of the samples from diarrhoeal children (n=136), and their respective demographic and clinical metadata, were input into a multinomial logistic regression model to identify factors explaining the grouping of the four CSTs (Table 6.2). CST2 was selected as a reference as it represented a mature gut microbiota configuration. I found that age, WAZ score, infection type (bacterial, viral, or mixed infection (Table 6.1)), and feeding practices remained significantly associated with CST classification ($p<0.05$; likelihood ratio test) (Table 6.3). Furthermore, younger age was significantly associated with CST1 (median age=16 months; $p=0.015$; two-tailed Z test) and CST3 (median age=15 months; $p=0.042$; two-tailed Z test) (Table 6.2), while the faecal samples in CST2 and CST4 were generally associated with older children (median age=22 and 23 months, respectively). This age differential may be explained by the maturation of the gastrointestinal microbiota in children as the *Bifidobacterium*-rich stage (CST1) precedes the *Bacteroides*-Firmicutes stage (CST2) (Dogra *et al.*, 2015). Additionally, I found that a lower WAZ score was associated with CST4 (median WAZ score=-1.1; $p=0.008$; two-tailed Z test), while CST3 contained a significantly greater proportion of faecal samples from children infected with bacterial diarrhoeal pathogens (73%; $p=0.001$; two-tailed Z test; Table 6.2). Lastly, children that were exclusively breastfed were significantly associated with CST1 ($p=0.015$; two-tailed Z test). However, there was no significant difference in feeding patterns according to age ($p=0.422$, Kruskal-Wallis rank sum test), suggesting that the association of breastfeeding with CST1 was not a product of clustering with younger age group.

Table 6.2 Demographic and clinical characteristics of 136 diarrhoeal individuals included in the multinomial regression analysis, stratified by the membership of bacterial community state type (CST). CST2 was selected as the reference group for regression. (*) Figures showing statistically significant differences compared to reference group.

	CST1 (*Bifidobacterium*) n (%)	**CST2 (*Bacteroides*) n (%)**	**CST3 (*Streptococcus*) n (%)**	**CST4 (*Escherichia*) n (%)**
Diarrhoeal patients' characteristics	N = 25	N = 33	N = 49	N = 29
Median age in months (range)	16 (4 - 30)*	22 (8 - 50)	15 (2 - 53)*	23 (5 - 59)
Male sex	15 (60.0)	15 (45.5)	32 (65.3)	15 (51.7)
Median weight-for-age Z score (range)	-0.53 (-3.42 - 1.24)	-0.27 (-2.40 - 1.78)	-0.5 (-2.81 - 2.39)	-1.1 (-3.17 - 2.22)*
Feeding behavior				
Breastfed only	18 (72.0)*	15 (45.5)	16 (32.7)	19 (65.5)
Breastfed and formula-milk fed	2 (8.0)	10 (30.3)	16 (32.7)	7 (24.1)
Formula-milk fed only	5 (20.0)	8 (24.2)	17 (34.7)	2 (10.3)
Urban residence	21 (84.0)	27 (81.8)	41 (83.7)	21 (72.4)
Income bracket (monthly)				
< 145 USD	5 (20.0)	8 (24.2)	13 (26.5)	10 (34.5)
145-483 USD	19 (76.0)	24 (72.7)	31 (63.3)	17 (58.6)
> 484 USD	1 (4.0)	1 (3.0)	5 (10.2)	2 (6.9)
With vomiting	12 (48.0)*	27 (81.8)	33 (67.3)	20 (69.0)
With dysentery	10 (40.0)	15 (45.5)	24 (49.0)	11 (37.9)
Infection aetiologies				
Viral mono infection	10 (40.0)	12 (36.4)	6 (12.2)	6 (20.7)
Bacterial mono infection	11 (44.0)	16 (48.5)	36 (73.5)*	16 (55.2)
Mixed infection	4 (16.0)	5 (15.2)	7 (14.3)	7 (24.1)

Table 6.3 Results of multinomial logistic regression for association of demographic and clinical predictors for classification of community state types (CSTs)

Patient characteristics	Odds ratio of associated factor (95% CI*); *p* value$		
	CST1 (*Bifidobacterium*#; N=25) *v* CST2	**CST3 (*Streptococcus*; N=49) *v* CST2**	**CST 4 (*Escherichia*; N=29) *v* CST2**
Age in months	0.92[0.86-0.98]; ***p*=0.015**	0.95[0.91-1.00]; ***p*=0.042**	0.99[0.94-1.04]; *p*=0.574
Male	2.14[0.62-7.41]; *p*=0.230	2.81[0.98-8.03]; *p*=0.054	2.12[0.67-6.65]; *p*=0.200
Weight-for-age Z score	0.63[0.35-1.12]; *p*=0.113	0.69[0.42-1.12]; *p*=0.135	0.46[0.26-0.81]; ***p*=0.008**
Breastfed and formula-milk fed (compared to breastfed-only)	0.10[0.01-0.64]; ***p*=0.015**	0.81[0.24-2.76]; *p*=0.737	0.44[0.11-1.71]; *p*=0.234
Formula-milk fed only (compared to breastfed-only)	0.26[0.06-1.21]; *p*=0.087	1.38[0.38-5.01]; *p*=0.621	0.25[0.05-1.36]; *p*=0.108
Urban residence (compared to rural)	1.60[0.29-8.95]; *p*=0.593	2.17[0.51-9.3]; *p*=0.297	1.5[0.33-6.8]; *p*=0.600
Monthly income 145-483 USD (compared to monthly income < 145 USD)	1.35[0.31-5.93]; *p*=0.690	0.53[0.16-1.79]; *p*=0.305	0.48[0.13-1.74]; *p*=0.262
Monthly income > 484 USD (compared to monthly income < 145 USD)	3.75[0.15-94.77]; *p*=0.422	3.28[0.26-41.97]; *p*=0.360	2.42[0.14-41.78]; *p*=0.544
Vomiting	0.19[0.05-0.77]; ***p*=0.020**	0.68[0.20-2.26]; *p*=0.525	0.42[0.11-1.61]; *p*=0.203
Dysentery	0.63[0.17-2.35]; *p*=0.496	0.76[0.26-2.24]; *p*=0.624	0.56[0.17-1.86]; *p*=0.342
Bacterial infection (compared to viral)	2.12[0.43-10.52]; *p*=0.358	13.61[3.01-61.52]; ***p*=0.001**	4.02[0.81-20.04]; *p*=0.089
Mixed infection (compared to viral)	1.21[0.19-7.78]; *p*=0.844	5.07[0.9-28.58]; *p*=0.066	3.31[0.59-18.74]; *p*=0.175

Figures in bold indicate $p<0.05$. *CI: Confidence Interval, $ *p* value calculated through two tailed Z test, # Bacterial taxon which is predominant in each CST

6.3.4 Changes in taxonomic abundance in the faeces of children with diarrhoea

Streptococcus and *Escherichia* species were highly abundant in the diarrhoeal gut communities associated with CST3 and CST4, respectively. However, 64/145 (44%) of the faecal bacterial community structures from diarrhoeal patients were more similar to those from the control groups (CSTs 1 and 2). I used DESeq2 to identify the OTUs that were differentially abundant in the diarrhoeal samples of all four CSTs, taking into account the overarching discrepancy in composition due to different CST membership, aiming to minimize effects associated with confounding factors such as age, WAZ scores, and infection type. Seventeen OTUs were found to be at least four-fold higher in abundance in diarrhoeal samples than control samples ($p<0.05$); the majority of these were *Escherichia* species or organisms associated with the oral cavity, including *Fusobacterium*, *Gemella,* and *Actinomyces* (Fig 6.8A). The most significant difference in a bacterial species in diarrhoeal faecal samples was associated with *Fusobacterium mortiferum*, which was 2^{10} fold more abundant in diarrhoeal faecal samples than control samples ($p<0.0001$). Furthermore, *Sutterella*, *Megamonas*, and *Enterococcus* species were also in higher abundance in diarrhoeal faecal samples in comparison to control samples. This analysis also identified 19 OTUs that were significantly depleted in diarrhoeal samples ($p<0.05$). The majority of these OTUs (13/19) were gut colonizers belonging to the orders Clostridiales and Erysipelotrichales, including *Subdoligranulum*, *Roseburia*, *Eubacterium*, *Coprococcus*, *Catenibacterium,* and Ruminococcaceae. Among these, the most depleted taxon was the short chain fatty acid (SCFA) producer *Blautia hansenii*, which was significantly reduced in samples from those with diarrhoea, regardless of the overall community group of the sample ($p<0.0001$) (Fig 6.9). *B.*

bifidum, *B. uniformis,* and *L. reuteri* were also significantly associated with a healthy microbial configuration, albeit with different proportions in the various CSTs, likely reflecting age-specific development of the gut microbiota (Fig 6.9). In addition, I performed the same analysis on a subset of samples belonging to CSTs 1 and 2, and this produced comparable results to the aforementioned.

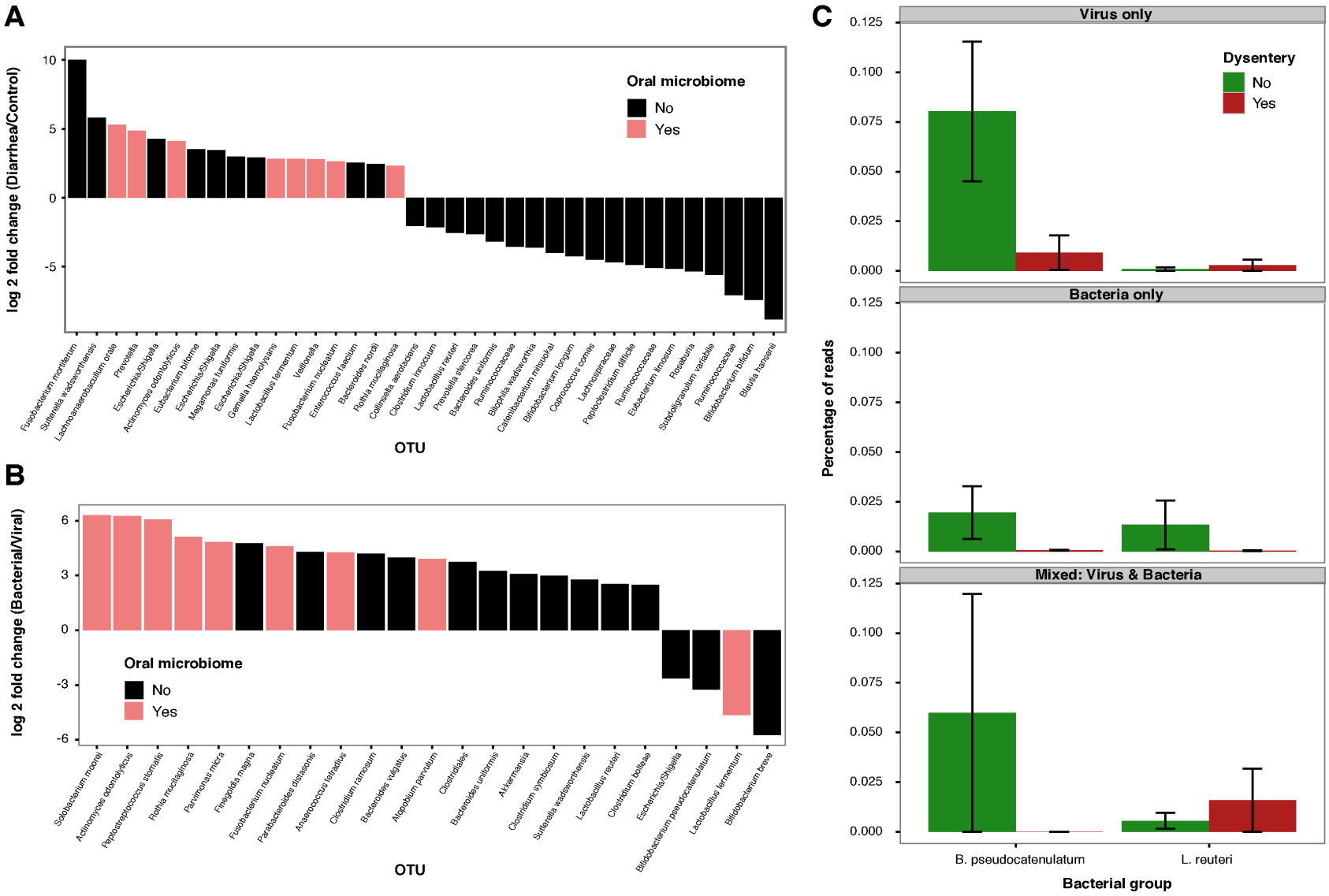

A
log 2 fold change (Diarrhea/Control)
10
5
0
-5
Oral microbiome
No
Yes
Fusobacterium mortiferum
Sutterella wadsworthensis
Lachnoanaerobaculum orale
Prevotella
Escherichia/Shigella
Actinomyces odontolyticus
Eubacterium biforme
Escherichia/Shigella
Megamonas funiformis
Escherichia/Shigella
Gemella haemolysans
Lactobacillus fermentum
Veillonella
Fusobacterium nucleatum
Enterococcus faecium
Bacteroides nordii
Rothia mucilaginosa
Collinsella aerofaciens
Clostridium innocuum
Lactobacillus reuteri
Prevotella stercorea
Bacteroides uniformis
Ruminococcaceae
Blautia wadsworthia
Catenibacterium mitsuokai
Bifidobacterium longum
Coprococcus comes
Lachnospiraceae
Peptostreptococcaceae
Ruminococcus difficile
Eubacterium limosum
Roseburia
Subdoligranulum variabile
Ruminococcaceae
Bifidobacterium bifidum
Blautia hansenii
OTU
B
log 2 fold change (Bacterial/Viral)
6
3
0
-3
-6
Oral microbiome
No
Yes
Solobacterium moorei
Actinomyces odontolyticus
Peptostreptococcus stomatis
Rothia mucilaginosa
Parvimonas micra
Finegoldia magna
Fusobacterium nucleatum
Parabacteroides distasonis
Anaerococcus lactolyticus
Clostridium ramosum
Bacteroides vulgatus
Atopobium parvulum
Clostridiales
Bacteroides uniformis
Akkermansia
Clostridium symbiosum
Sutterella wadsworthensis
Lactobacillus reuteri
Clostridium bolteae
Escherichia/Shigella
Bifidobacterium pseudocatenulatum
Lactobacillus fermentum
Bifidobacterium breve
OTU
C
Virus only
Bacteria only
Mixed: Virus & Bacteria
Dysentery
No
Yes
Percentage of reads
0.125
0.100
0.075
0.050
0.025
0.000
B. pseudocatenulatum
L. reuteri
Bacterial group

Figure 6.8. Bacterial taxa showing significantly different abundance among the examined classes

OTUs were identified to be of significantly differential abundance between groups in examination, as detected and filtered by DESeq2. In short, only OTUs with adjusted *p* values <0.05, estimated fold change >4 or <1/4, and estimated base mean >30 were considered significantly differentially abundant and included in the plot. (A) Ratio of the log2 fold change of OTUs that differ between diarrhoeal and control stools, accounting for the different general microbial configurations in four CSTs (N=199). (B) Ratio of the log2 fold change of OTUs that differ between bacterial and viral diarrhoeal infections, accounting for the different general microbial configurations in four CSTs (N=119). (C) Relative abundances of OTUs that differ between dysenteric and non-dysenteric stools, accounting for different types of infection (N=142).

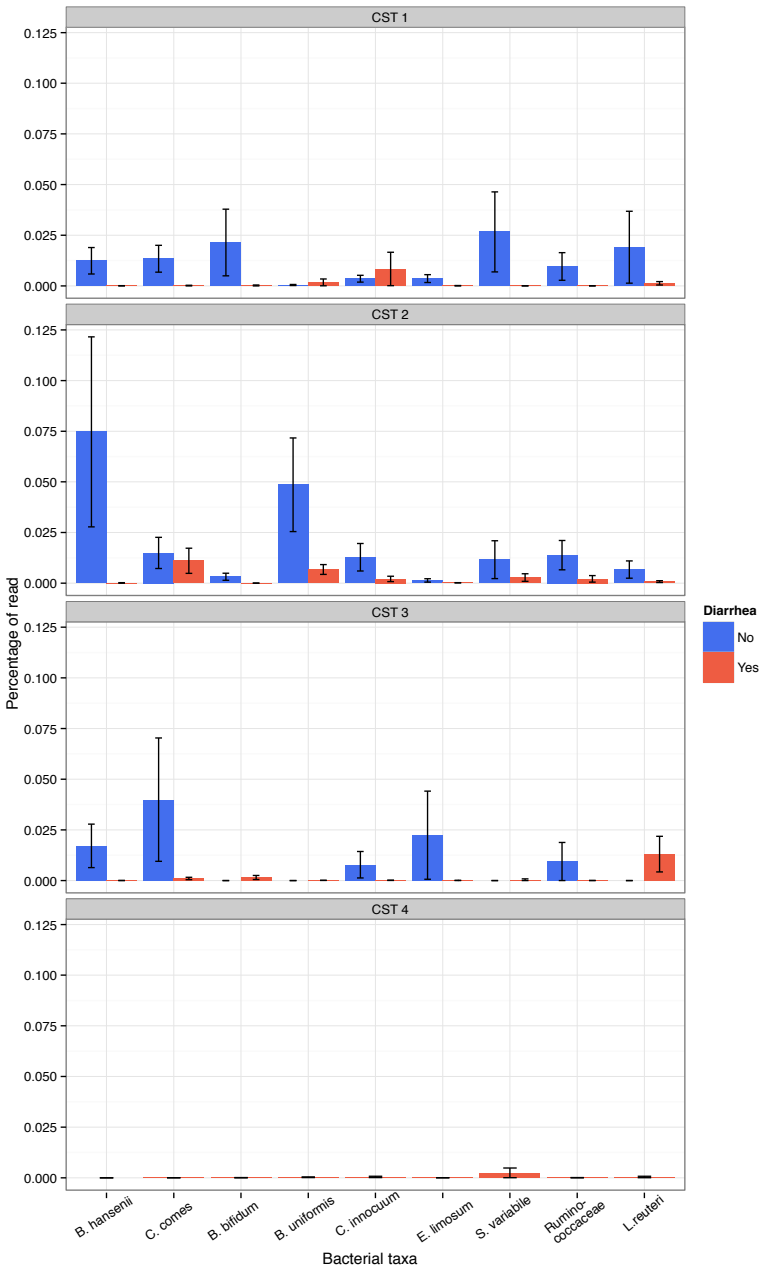


Figure 6.9 Bacterial taxa with decreased abundance during the early phase of diarrhoea The proportions of selected bacterial taxa, which were identified to be significantly depleted in diarrhoeal faecal samples across four CSTs (N=199) based on DESeq2 analysis (Fig 6.8), were shown for each CST. Bar plot shows the mean relative abundance of the bacterial taxa within each CST, with whiskers denoting the standard error.

I next constructed a correlation network that encompassed representative faecal taxa from both control and diarrhoeal samples. The resulting correlation network formed a tightly connecting interaction between members of the Clostridiales, Erysipelotrichales, and Bacteroidales, which indicated a positive interaction between these colonizers of a healthy gut (Fig 6.10). A further positive interacting cluster was composed of genera that are frequently associated with the human oral cavity, including *Fusobacterium nucleatum*, *Parvimonas micra*, *Peptostreptococcus stomatis*, *Solobacterium moorei*, *Gemmella haemolysans*, *Actinomyces odontolyticus*, *Lachnoanaerobaculum orale*, *Abiotrophia defectiva*, *Atopobium parvulum*, *Rothia mucilaginosa,* and *Streptococcus salivarius* (Fig 6.10). Conversely, *Rothia* was negatively correlated with *Bacteroides vulgatus* and *Faecalibacterium prausnitzii*, which are both indicators of a healthy and mature gastrointestinal microbiota (Subramanian *et al.*, 2014; Bäckhed *et al.*, 2015). I additionally found that *Fusobacterium mortiferum*, the significantly enriched taxon in faecal samples from diarrhoeal children, exhibited a negative association with a key component of the gut Clostridiales network (*Clostridium innocuum*). This suggests that the colonization of *F. mortiferum* may inhibit the proliferation of related dependent taxa such as *Blautia*, *Ruminococcus,* and several *Bacteroides*. The negative correlation observed between *Bifidobacterium longum* subsp. *infantis* and *Bifidobacterium breve* likely reflects a competitive ecology as the gut microbiota matures (Avershina *et al.*, 2013).

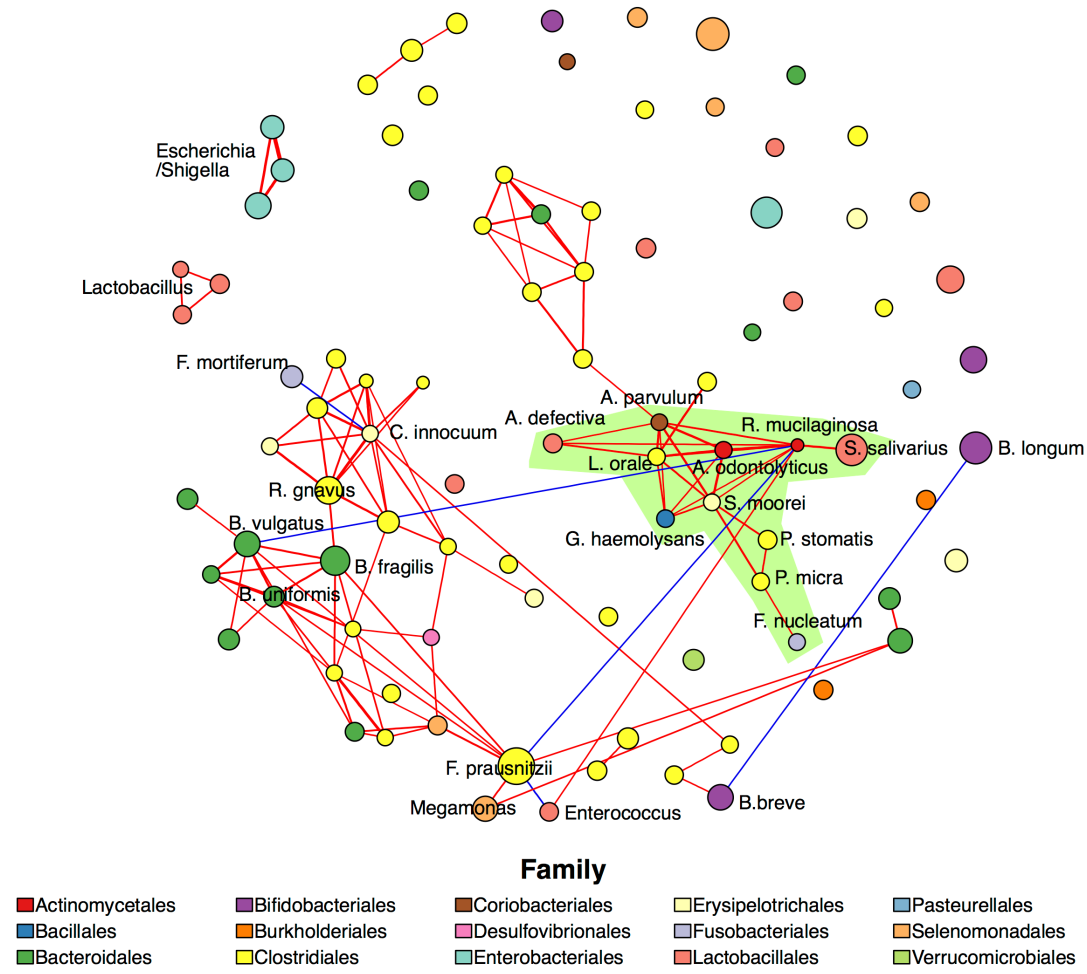


Figure 6.10 Correlation network of the healthy and diarrhoeal microbiome

Figure shows the correlation network of the 92 most represented OTUs sampled from 199 stool isolates, defined as OTUs with occurrence in at least 10 samples, constructed using the SparCC. Only correlations with calculated *p* value ≤0.05 and absolute magnitude ≥0.25 were shown in the network. Positive and negative interactions were denoted as a red and blue solid line respectively, with line weight proportional to correlation strength. The OTUs (nodes) were coloured based on taxonomic family (see legend), with sizes proportional to their relative abundances. The light green shaded area covers OTUs identified as members of normal human oral microbiota (through comparison with the Human Oral Microbiome Database).

6.3.5 Taxonomic changes associated with bacterial induced diarrhoea and dysentery

In order to explore the differential composition of the gut microbiota in response to the different aetiological agents (viral vs. bacterial), I examined a subset of 119 diarrhoeal faecal samples (viral infections=39; bacterial infections=80) to eliminate potential misclassification induced by 23 cases of mixed infections. Analysis using DESeq2 identified a group of 19 OTUs that were significantly more abundant in bacterial infections across the four CSTs. Among these, eight taxa were more commonly associated with human oral microbiota (Fig 6.8B); *Actinomyces* and *Rothia* were the most prevalent (assessed by the DESeq2 calculated base mean value). Other organisms with high relative proportion in the faecal samples from children with bacterial induced diarrhoea included *C. ramosum* and *B. vulgatus*. Conversely, four OTUs were significantly elevated in the viral diarrhoeal infections; these included the *Bifidobacterium* species, *B. breve* and *B. pseudocatenulatum* (Fig 6.8B).

I lastly categorized 142 diarrhoeal faecal samples based on the diagnosis of dysentery (the visible presence of mucoid and/or blood in the stool; n=60), accounting for the differential microbiota changes due to types of infectious agents (bacterial, viral, or mixed infection). I identified two OTUs that were significantly associated with faecal samples from non-dysenteric diarrhoea (DESeq2, $p<0.05$) (Fig 6.8C). *B. pseudocatenulatum* was almost 2^5 fold less abundant in dysenteric faeces in comparison to non-dysenteric diarrhoeal faeces; this was consistent across all of the corresponding aetiological agents (Fig 6.8C). For bacterial infections, *L. reuteri* was uniformly more

depleted in the examined dysenteric faecal samples, but its variability was higher in dysenteric stools in those with a viral or mixed infection. When I accounted only for the stool with visible blood (n=10), the analysis revealed that *S. gallolyticus* was the most depleted taxon in dysentery compared to watery diarrhoea (DESeq2 Wald test without effect of infection types, $p<0.05$). Other taxa of decreased abundance included *Escherichia*, *B. pseudocatenulatum*, *Enterococcus* and several oral microbiome species. However, these interpretations are to be confirmed in larger studies.

6.4 Discussion

Here I assessed the impact of diarrhoeal infection on the composition of the gut microbiota in Vietnamese children. My results predict that the gut microbiota in young children exhibits differential responses during the early phase of diarrhoea, which can be grouped into microbial community structures that either closely resemble or are highly divergent from those of healthy children. Approximately 44% of the diarrhoeal cases in this study retained a *Bifidobacterium*- or *Bacteroides*-rich structure. This association largely resembles the well-studied age-dependent normal gut microbiota in children (Dogra *et al.*, 2015), outlining a lack of global dysbiosis in these children. Among the various perturbed microbial states, I found that age, nutritional status, breastfeeding, and diarrhoeal aetiology contributed to the composition of the bacterial communities during the early phase of diarrhoea. However, since the WAZ score was collected after the onset of diarrhoea, weight could be subjected to temporary fluctuation and its surrogate for nutritional status should be interpreted with caution. A recent study on diarrhoeal microbiome from U.S patients reported that diarrhoeal microbial structures could be

grouped into four major clusters irrespective to the aetiologies, of which two are almost exclusively restricted to diarrhoeal patients and associated with higher prevalence of *Escherichia* (Singh *et al.*, 2015). In my study, *Streptococcus* was associated with younger age and bacterial infection, whilst *Escherichia* was associated with older children and a poor nutritional status. Indeed, analysis of the microbiota in diarrhoeal samples from young children (<2 years old) predominantly infected with pathogenic *E. coli* revealed that *Streptococcus* (*S. gallolyticus* and *S. salivarius*) was the most abundant organisms during the early phase of diarrhoea (Sarker *et al.*, 2016). These two species were also dominant in my study. In contrast, *Escherichia* were overrepresented in the faecal samples from children (2–3 years old) infected with cholera in Bangladesh (Monira *et al.*, 2013). Due to limitations of the study, I was unable to assess how these initial differential microbial communities affect diarrhoeal severity and recovery. Dysentery and antimicrobial treatment may further confound these associations. However, the ratio of *Streptococcus*/*Bifidobacterium* showed a significant positive correlation with diarrhoeal faecal output and duration of hospitalization, suggesting that the higher degree of divergence from the healthy microbiota may predict a more severe clinical presentation (Sarker *et al.*, 2016).

I additionally observed that obligate anaerobes belonging to the orders Clostridiales and Erysipelotrichales were consistently depleted in the diarrhoeal cases in comparison to the controls. This group included multiple SCFA producers including *Blautia*, *Subdoligranulum*, *Roseburia,* and *Eubacterium*. This finding may explain the transient depletion of SCFA following diarrhoea, resulting in poor water and electrolyte absorption

and a shortage of metabolic energy for enterocytes, which may lead to dehydration and further complications (Monira *et al.*, 2010). These data are in accordance with previous studies, which have shown that Clostridiales become gradually enriched during the diarrhoeal recovery period (Monira *et al.*, 2013; Hsiao *et al.*, 2014; David *et al.*, 2015). This loss of obligate anaerobes is coupled with the proliferation of facultative anaerobes, such as the *Streptococcus* and *Escherichia* species (Pop *et al.*, 2014; David *et al.*, 2015). Unpredictably, I observed that a substantial component of the microbiota from the human oral cavity was consistently represented in the diarrhoeal faecal samples. I propose that the transiently oxygenated environment and potential decrease in bacterial competition in the diarrhoeal gut enhances the colonization potential of these species (David *et al.*, 2015). This may explain why *S. salivarius*, a dominant taxon of the oral cavity of young children, is capable of establishing an exceptionally high rate of colonization in suckling infants. Oral commensal organisms, predominantly *S. salivarius* and *F. nucleatum*, may serve as nucleators for the adherence of other bacteria, eventually forming polymicrobial biofilms that are capable of surviving mechanical and chemical stresses during their passage through the large intestine (Bradshaw *et al.*, 1998; Levesque, Lamothe and Frenette, 2003). Alternatively, the diarrhoea-associated obligate anaerobe, *F. mortiferum*, which was found not to be associated with oral organisms, exhibited antagonism with the Clostridiales. This finding is in agreement with the spike of *Fusobacterium* observed in early phase diarrhoea (Hsiao *et al.*, 2014; Becker-dreps *et al.*, 2015). Unlike the majority of *Fusobacterium*, *F. mortiferum* is extremely pleomorphic and bile resistant, with the ability to rapidly accumulate and metabolize a broad spectrum of sugars independent of amino acid fermentation (Robrish, Oliver and Thompson, 1991; Krieg *et al.*, 2010; Gupta

and Bhandari, 2014). Furthermore, poultry isolated *F. mortiferum* has been shown to secrete a bacteriocin that can inhibit the growth of Gram-positive *Bacillus* species (Portrait, Cottenceau and Pons, 2000). Taken together, this evidence implies that *F. mortiferum* is more resilient to environmental stresses and able to occupy the anaerobic niche of the gut, possibly accelerating the depletion of Clostridiales via bacteriocin-mediated competition.

Dysentery denotes a more severe form of diarrhoea, frequently requiring antimicrobial treatment and resulting in longer periods of hospitalization (Traa *et al.*, 2010; Thompson *et al.*, 2015). To date, only *Lactobacillus ruminis* has been found to be depleted in dysenteric faecal samples (Pop *et al.*, 2014). Here, I found that *L. reuteri* was significantly diminished in dysenteric diarrhoea; I speculate that this discrepancy in species level identification may be induced by variation in environmental and genetic factors of the studied populations. In addition, I report here that a *Bifidobacterium* species, *B. pseudocatenulatum*, is substantially depleted in faecal specimens following bacterial infection; this effect is amplified during dysentery. Previous research examining the V1-V2 region of the 16S rRNA failed to efficiently identify *Bifidobacterium*, thus discounting its contribution (Pop *et al.*, 2014). The oral administration of live *B. pseudocatenulatum* in induced cirrhotic and diabetic mice has been shown to restore the gut integrity and initiate an anti-inflammatory cytokine profile in the intestine, leading to a diminished systematic inflammatory response (Moratalla *et al.*, 2014, 2016; Moya-Pérez, Neef and Sanz, 2015). Further, this potential probiotic also induces monocyte-derived macrophages to undergo transition into the anti-inflammatory M2 phase

(Moratalla *et al.*, 2015). It has been shown that the colonization of *B. pseudocatenulatum* or *B. breve*, but not other *Bifidobacterium*, protect mice from lethal Shiga toxin producing *E. coli* (STEC) infections (Asahara *et al.*, 2004). Treatment for dysentery is currently focused on antimicrobial therapy, and the usage of broad spectrum antimicrobials quickly depletes the balanced microbiota derived metabolites such as SCFAs, leading to disruption of the gut immune homeostasis (Lopez *et al.*, 2014). Therefore, I propose that *B. pseudocatenulatum* may be a potential dysentery-oriented probiotic candidate to reduce inflammation associated pathological conditions and accelerate the recovery of the microbiota to a healthy state.

Some limitations should be considered in the context of my investigation. First, the EMIRGE mediated read recovery rate was consistently lower in diarrhoeal samples. This means that a proportion of uncharacterized taxa were unavailable for examination, and it may impact on the true estimation of the diversity within these samples. Second, an exhaustive search for other diarrhoeagenic aetiologies by molecular methods was not possible during this study; this may underestimate the true incidence of mixed infections. I propose that more comprehensive approaches, such as the Luminex xTAG multi-pathogen panel assay, could be incorporated in further studies to improve diagnostic accuracy (Duong *et al.*, 2016). Third, the non-diarrhoea controls included are not ideally representative of the healthy population, given that some of these children sought medical help for nutritional concerns. Notwithstanding these limitations, through sequencing an extended region of the 16S rRNA gene, it was possible to obtain high confidence taxonomic assignment up to the species level, portraying the highly complex gut

microbiota in greater resolution. The inclusion of various diarrhoeal aetiologies and the statistical methods employed here further permitted us to reveal the general impact of infectious diarrhoea on the gut microbiota. Diarrhoea cases with unknown aetiologies were excluded from this study, and this may underrepresent the true heterogeneity of diarrhoeal causes and their differing influences on the microbiota. Regarding the correlation network analysis, it is apparent that SparCC detects fewer negative (competitive relationships) than positive correlations (mutualism and commensalism). Unlike positive correlations, which are more readily associated with the co-occurrence of OTUs, negative correlations are typically inferred by the mutual exclusions of OTUs. SparCC has been shown to possess increased precision in detecting negative correlations in comparison to other methods, but correlation detection is limited to those with greater power and low noise (Weiss *et al.*, 2016).

In conclusion, by studying the microbiota of children with diarrhoea associated with a diverse range of confirmed aetiological agents, I have generated a comprehensive insight into the effects of infectious diarrhoea on the gut microbiota. Therefore, this study provides a pivotal understanding of the impact of infectious diarrhoea on the gut microbiota of children, particularly ones in low-middle income urban settings in Southeast Asia. Future work, through longitudinal study designs and the employment of shotgun metagenomics, should address how the initial differential responses in gut bacterial composition may impact disease progression, such as diarrhoea clearance and recovery rate of the depleted microbiota.

7 General Discussion

Despite improvements in clean water accessibility, sanitation and treatment, *Shigella* remains an important aetiological agent of paediatric diarrhoea in tropical and developing settings. State-of-the-art genomic analysis has been applied successfully to gain insights into the evolutionary history and phylogenetic structure of different *Shigella* species, including *S. sonnei* (Holt *et al.*, 2012), *S. flexneri* (Connor *et al.*, 2015), and *S. dysenteriae* serotype 1 (Njamkepo *et al.*, 2016). Nevertheless, these studies provided insufficient resolution in light of rapidly arising and clinically important variants, such as FQ resistant *Shigella*, and do not provide any detailed epidemiological history of an endemic region. This thesis aimed to bridge these knowledge gaps and further our understanding of the evolution and dynamics of this pathogen.

7.1 The surge of fluoroquinolone resistant *Shigella sonnei*

The data presented in Chapters 3 and 4 was the first attempt to characterize FQ resistant *S. sonnei* at a genomic level. I concluded that extant FQ resistant *S. sonnei* are clonal and likely evolved and spread in South Asia before disseminating intercontinentally. Regional diversifications soon followed, creating one clone circulating in Southeast Asia and another appearing to instigate sustained transmission within Europe. Reductive evolution, most notably via deletion of the flagellum encoding operon, and adaptation under oxidative stress both highlighted the evolution of this global FQ resistant clone. Though I demonstrated that FQ resistant *S. sonnei* has been successful in its global propagation, its clinical interpretation is still debatable. I found that there was a decrease in plasma IL1-Ra titre for patients infected with FQ resistant *S. sonnei*, but how this may translate to

patients' inflammatory status and disease severity requires further investigation. In addition, antimicrobial resistance does not always result in treatment failure, especially for self-limiting infections like shigellosis. A recent study on paediatric diarrhoea in Vietnam found that hospitalization length for *Shigella* infected patients is similar regardless of the FQ susceptibility profile of the associated organism (Duong *et al.*, 2018). Furthermore, antimicrobial treatment does not improve, and may worsen, clinical outcomes in this studied population. Thus, FQ resistance likely poses a more significant threat to certain high-risk cohorts, including the malnourished, the immunocompromised, and the elderly.

FQ resistance has previously been reported in other *Shigella* species. Both epidemiological surveillance and genomic analysis pointed to the emergence and abrupt decline of FQ resistant *S. dysenteriae* serotype 1 during 2002-2003, of which circulation was restricted to South Asia (Bhattacharya *et al.*, 2003; Naheed *et al.*, 2004; Njamkepo *et al.*, 2016). FQ resistant *S. flexneri* was detected as early as 2001 in China, and its presence became both geographically and serotypically widespread in the country (Wang *et al.*, 2005; Pu *et al.*, 2009; Qin *et al.*, 2017). Compared to these two species, FQ resistance emerged in *S. sonnei* much later in 2007. However, the first contributing *gyrA*-S83L mutation likely occurred pre-1996, nine years before the fixation of the subsequent *parC*-S80I mutation. The reason underlying this temporal distance between these two events is uncertain, and may be related to the low circulation of *S. sonnei* in South Asia during this period (von Seidlein *et al.*, 2006; Taneja, 2007; Ud-Din *et al.*, 2013). Recent experimental and modelling work has suggested that the *parC*-S80I mutation is the key

determining factor in the stepwise evolution of resistance against FQ (Huseby *et al.*, 2017). This mutation occurs much less frequently during evolution but confers a significant increase in FQ MIC and potentially a non-inferior fitness of the bacterium. It is proposed that *parC*-S80I is favourably selected in abundance of mutation supply, fulfilled either by a large population size or a high mutation rate, when antimicrobial pressure is high. Given that the mutation rate of *S. sonnei* is relatively stable, the first scenario appears to be more plausible (Holt *et al.*, 2012, 2013). High human population densities in South Asia could promote extensive and sustained *Shigella* transmission, resulting in a large bacterial population. Suboptimal public health measures in the region, exemplified by the fact that only 40% of the Indian population has access to improved sanitation (World Bank, 2015), may further amplify the transmission cycle of *Shigella* in this population. This expansion has arisen against a backdrop of rapidly increasing FQ use for treatment of multiple enteric and febrile diseases since the turn of this century. These contributing factors might render South Asia a unique focal point for the emergence of human restricted FQ resistant enteric bacteria, including *Shigella* and *Salmonella* Typhi (Wong *et al.*, 2015).

It has been experimentally demonstrated that QRDR mutations, similar to those acquired by FQ resistant *S. sonnei* in this study, may carry no detrimental or limited fitness cost to the resistant Enterobacteriaceae, even in the absence of FQ pressure (Marcusson, Frimodt-Møller and Hughes, 2009; Baker *et al.*, 2013). Alternatively, FQ resistance has been coupled with the successful clonal propagations of several multi-drug resistant pathogens, including *Staphylococcus aureus*, *Klebsiella pneumoniae*, *E. coli*, *Clostridium*

difficile and *Neisseria gonorrhoeae* (Grad *et al.*, 2016; Fuzi, Szabo and Csercsik, 2017). All these major FQ resistant clones were found to harbour specific QRDR mutations combinations, indicating that these resistance genotypes induce a minimal fitness disadvantage. Although little is known about the impact of alleviating FQ pressure on the clonal dominance of FQ resistant bacteria in nature, it is speculated that withdrawal of such pressure in clinical settings is unlikely to discontinue the dominance of FQ resistant *S. sonnei* in the transmission chain. However, future research is warranted to challenge this hypothesis, as well as to develop the best practices for controlling and treating new emerging antimicrobial resistant clones.

7.2 Historical epidemiology of *Shigella* in Southeast Asia

The large-scale analyses discussed in Chapter 5 revealed a highly dynamic population succession of both *S. flexneri* and *S. sonnei* in continental Southeast Asia. The majority of extant *Shigella* isolates are progenies of relatively recent introductions, mostly occurring in the 1980s-1990s. Geographical connections facilitate the propagation of these pathogens across the region, forming local or supra-regional expansions. The most striking finding lays in the regularity of clonal replacement within the genus, observed in both major *S. flexneri* serotypes and *S. sonnei*, with these ecological successions often coupled with incoming resistance to contemporaneously used antimicrobials. These results resonate with those recently deduced for another enteric pathogen, *Vibrio cholerae*. A single bacterial lineage responsible for the seventh pandemic cholera was shown to have been introduced into Africa on twelve distinct occasions, with the later and more resistant waves replacing the pre-existing antimicrobial susceptible populations

(Weill *et al.*, 2017). Furthermore, these introductions only began since the 1970s, and the authors argued that intense human migration and contact during the few decades of globalization resulted in inter and intra-continental cholera transmission. This finding also downplays the role of aquatic *V. cholerae* in the pandemic's transmission chain. I speculate that although species diversity in both *S. flexneri* and *V. cholerae* are sustained via occupation of respective environmental niches, the majority of existent clinical burden could be attributed to expansions of certain clones, probably owing to increased human mobility and contact.

Future research should focus on many unanswered questions of *Shigella* epidemiology. Particularly, the rise of FQ resistant *S. flexneri* has not been deciphered on the genomic and phylogenetic scale. Collaborative efforts are ideal to include data from other less well-studied endemic regions, such as China, South America, and Africa. Much emphasis should be paid on how to translate the findings from molecular epidemiology studies into actionable plans, to enhance public health and clinical managements of shigellosis.

7.3 Microbiota perturbation under infectious diarrhoea

I dedicated Chapter 6 to explore how the gut microbiome is modified in response to infectious diarrhoea in Vietnamese children. I found that diarrhoea could generate a faecal microbiome that either generally resembles (i.e. *Bifidobacterium* or *Bacteroides* rich) or is highly divergent (i.e. *Streptococcus* or *Escherichia* rich) from the healthy state. Consistent increases in abundances of *F. mortiferum*, *Escherichia*, and oral commensals were shown for all of these diarrhoeal microbial configurations when compared to non-

diarrhoeal controls, proposing similar mechanistic interactions. However, my study was limited to the sole description of microbial communities during the early phase of diarrhoea, so the findings cannot yet offer significant clinical applications. Future research should employ a longitudinal design in order to document the microbiota alternations from diarrhoeal onset to recovery. It is of considerable interest to evaluate how different initial microbiota signatures affect key clinical outcome, such as diarrhoea clearance time. This might allow us to identify taxonomical groups critical for expediting recovery, forming a platform for evidence-based diarrhoeal probiotics. In addition, sequence-based microbiota analyses should be complemented with experimental rigour to confirm any inferred causality or reveal mechanistic interactions, as has been conducted previously in the search for a probiotic targeting *C. difficile* infections (Buffie *et al.*, 2014). I acknowledge that the results may only be generalized to infectious diarrhoea in young children (< 5 years), and differing responses could be observed in older children or adults or in cases of non-infectious diarrhoea. Indeed, a recent longitudinal study with a dense sampling strategy revealed that induced iso-osmotic diarrhoea (IIOD) in adults led to transient disturbances, reducing the abundance of *Ruminococcus* in faecal samples immediately after diarrhoea (Fukuyama *et al.*, 2017). However, the faecal microbiome was assessed to return within days to its original, pre-intervention state. This indicates that the physical manifestation of diarrhoea may only target a specific and vulnerable bacterial group, and other dysbiosis that I observed may result from additional underlying mechanisms such as inflammation induced changes in immune response and oxygen gradient (Velly, Britton and Preidis, 2017).

8 References

- Van Acker, H. and Coenye, T. (2017) 'The Role of Reactive Oxygen Species in Antibiotic-Mediated Killing of Bacteria', *Trends in Microbiology*. Elsevier Ltd, xx, pp. 1–11. doi: 10.1016/j.tim.2016.12.008.
- Afroze, F., Ahmed, T., Sarmin, M., Shahid, A. S., Shahunja, K. M., Shahrin, L. and Chisti, M. J. (2017) 'Risk factors and outcome of Shigella encephalopathy in Bangladeshi children', *PLOS Neglected Tropical Diseases*, 138, pp. 1–11.
- Ahmed, Z. U., Sarker, M. R. and Sack, D. A. (1988) 'Nutritional requirements of shigellae for growth in a minimal medium', *Infection and Immunity*, 56(4), pp. 1007–1009.
- Al-Hasani, K., Henderson, I. R., Sakellaris, H., Rajakumar, K., Grant, T., Nataro, J. P., Robins-Browne, R. and Adler, B. (2000) 'The sigA Gene Which Is Borne on the she Pathogenicity Island of Shigella flexneri 2a Encodes an Exported Cytopathic Protease Involved in Intestinal Fluid Accumulation', *Infection and Immunity*, 68(5), pp. 2457–2463. doi: 10.1128/IAI.68.5.2457-2463.2000.
- Alizadeh-Hesar, M., Bakhshi, B. and Najar-Peerayeh, S. (2015) 'Clonal dissemination of a single Shigella sonnei strain among Iranian children during Fall 2012 in Tehran, I.R. Iran', *Infection, Genetics and Evolution*. Elsevier B.V., (June). doi: 10.1016/j.meegid.2015.06.024.
- Allison, G. E. and Verma, N. K. (2000) 'Serotype-converting bacteriophages and O-antigen modification in Shigella flexneri', *Trends in microbiology*. 2000/01/19, 8(1), pp. 17–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10637639.
- Ariza, R. R., Li, Z., Ringstad, N. and Demple, B. (1995) 'Activation of multiple antibiotic resistance and binding of stress-inducible promoters by Escherichia coli Rob protein.', *Journal of bacteriology*, 177(7), pp. 1655–1661.
- Asahara, T., Shimizu, K., Nomoto, K., Hamabata, T., Ozawa, A. and Takeda, Y. (2004) 'Probiotic Bifidobacteria Protect Mice from Lethal Infection with Shiga Toxin-Producing Escherichia coli O157 : H7', *Infect Immun*, 72(4), pp. 2240–2247. doi: 10.1128/IAI.72.4.2240.
- Assefa, S., Keane, T. M., Otto, T. D., Newbold, C. and Berriman, M. (2009)

'ABACAS: Algorithm-based automatic contiguation of assembled sequences', *Bioinformatics*, 25(15), pp. 1968–1969. doi: 10.1093/bioinformatics/btp347.

- Avershina, E., Storrø, O., Øien, T., Johnsen, R., Wilson, R., Egeland, T. and Rudia, K. (2013) 'Bifidobacterial succession and correlation networks in a large unselected cohort of mothers and their children', *Applied and Environmental Microbiology*, 79(2), pp. 497–507. doi: 10.1128/AEM.02359-12.
- Azmuda, N., Rahman, M. Z., Sultana, M., Jenssen, E. L., Khan, S. I. and Birkeland, N. K. (2012) 'Evidence of interspecies O antigen gene cluster transfer between Shigella boydii 15 and Escherichia fergusonii', *Apmis*, 120(12), pp. 959–966. doi: 10.1111/j.1600-0463.2012.02926.x.
- Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., Khan, M. T., Zhang, J., Li, J., Xiao, L., Al-Aama, J., Zhang, D., Lee, Y. S., Kotowska, D., Colding, C., Tremaroli, V., Yin, Y., Bergman, S., Xu, X., Madsen, L., Kristiansen, K., Dahlgren, J. and Jun, W. (2015) 'Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life', *Cell Host & Microbe*, 17(5), pp. 690–703. doi: 10.1016/j.chom.2015.04.004.
- Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M. A. and Alekseyenko, A. V. (2012) 'Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty', *Molecular Biology and Evolution*, 29(9), pp. 2157–2167. doi: 10.1093/molbev/msl161.
- Baele, G., Li, W. L. S., Drummond, A. J., Suchard, M. a. and Lemey, P. (2013) 'Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics', *Molecular Biology and Evolution*, 30(2), pp. 239–243. doi: 10.1093/molbev/mss243.
- Baker, K. S., Campos, J., Pichel, M., Della Gaspera, A., Duarte-Martínez, F., Campos-Chacón, E., Bolaños-Acuña, H. M., Guzman-Verri, C., Mather, A. E., Velasco, S. D., Zamudio Rojas, M. L., Forbester, J., Connor, T. R., Keddy, K. H., Smith, A. M., Lopez de Delgado, E. A., Angiolillo, G., Cuaical, N., Fernandez, J., Aguayo, C., Aguilar, M. M., Valenzuela, C., Morales Medrano, A. J., Esteve, A. S., Gustafson, N. W., Diaz Guevara, P. L., Montaño, L. A., Perez, E. and Thomson, N. R. (2017) 'Whole genome sequencing of Shigella sonnei through PulseNet Latin America and Caribbean: advancing global surveillance of foodborne illnesses',

Clinical Microbiology and Infection, pp. 1–9. doi: 10.1016/j.cmi.2017.03.021.

- Baker, K. S., Dallman, T. J., Ashton, P. M., Day, M., Hughes, G., Crook, P. D., Gilbart, V. L., Zittermann, S., Allen, V. G., Howden, B. P., Tomita, T., Valcanis, M., Harris, S. R., Connor, T. R., Sintchenko, V., Howard, P., Brown, J. D., Petty, N. K., Gouali, M., Thanh, D. P., Keddy, K. H., Smith, A. M., Talukder, K. a, Faruque, S. M., Parkhill, J., Baker, S., Weill, F.-X., Jenkins, C. and Thomson, N. R. (2015) 'Intercontinental dissemination of azithromycin-resistant shigellosis through sexual transmission: a cross-sectional study', *The Lancet Infectious Diseases*, 3099(15), pp. 1–9. doi: 10.1016/S1473-3099(15)00002-X.
- Baker, S., Duy, P. T., Nga, T. V. T., Dung, T. T. N., Phat, V. V., Chau, T. T., Turner, a. K., Farrar, J. and Boni, M. F. (2013) 'Fitness benefits in fluoroquinolone-resistant Salmonella Typhi in the absence of antimicrobial pressure', *eLife*, 2, pp. e01229–e01229. doi: 10.7554/eLife.01229.
- Balbi, K. J., Rocha, E. P. C. and Feil, E. J. (2009) 'The temporal dynamics of slightly deleterious mutations in Escherichia coli and Shigella spp', *Molecular biology and evolution*. 2008/11/06, 26(2), pp. 345–355. doi: 10.1093/molbev/msn252.
- Bangtrakulnonth, A., Vieira, A. R., Lo, D. M. A., Wong, F., Pornreongwong, S., Pulsrikarn, C., Sawanpanyalert, P., Hendriksen, R. S. and Aarestrup, F. M. (2008) 'Shigella from Humans in Thailand During 1993 to 2006: Spatial-Time Trends in Species and Serotype Distribution', *Foodborne pathogens and disease*, 5(6), pp. 773–784.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. a., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. a. and Pevzner, P. a. (2012) 'SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing', *Journal of Computational Biology*, 19(5), pp. 455–477. doi: 10.1089/cmb.2012.0021.
- Barbagallo, M., Di Martino, M. L., Marcocci, L., Pietrangeli, P., De Carolis, E., Casalino, M., Colonna, B. and Prosseda, G. (2011) 'A new piece of the Shigella Pathogenicity puzzle: spermidine accumulation by silencing of the speG gene', *PloS one*. 2011/11/22, 6(11), p. e27226. doi: 10.1371/journal.pone.0027226.
- Barceloux, D. G. (2008) 'Shigella Species (Shiga Enterotoxins)', in *Medical*

Toxicology of Natural Substances. John Wiley & Sons, Inc., pp. 150–155. doi: 10.1002/9780470330319.ch20.

- Bardhan, P., Faruque, a S. G., Naheed, A. and Sack, D. a (2010) 'Decrease in shigellosis-related deaths without Shigella spp.-specific interventions, Asia.', *Emerging infectious diseases*. 2010/10/30, 16(11), pp. 1718–23. doi: 10.3201/eid1611.090934.
- Bastin, D. A., Lord, A. and Verma, N. K. (1997) 'Cloning and analysis of the glucosyl transferase gene encoding type I antigen in Shigella flexneri', *FEMS microbiology letters*. 1997/11/22, 156(1), pp. 133–139. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9368372.
- Becker-dreps, S., Allali, I., Monteagudo, A., Vilchez, S., Hudgens, M. G., Rogawski, E. T., Carroll, I. M., Zambrana, L. E., Espinoza, F. and Azcarate-peril, M. A. (2015) 'Gut Microbiome Composition in Young Nicaraguan Children during Diarrhea Episodes and Recovery', *Am. J. Trop. Med. Hyg.*, 93(6), pp. 1187–1193. doi: 10.4269/ajtmh.15-0322.
- van den Beld, M. J. C. and Reubsaet, F. a G. (2012) 'Differentiation between Shigella, enteroinvasive Escherichia coli (EIEC) and noninvasive Escherichia coli', *European Journal of Clinical Microbiology and Infectious Diseases*. 2011/09/09, 31(6), pp. 899–904. doi: 10.1007/s10096-011-1395-7.
- Bennik, M. H. J., Pomposiello, P. J., Thorne, D. F. and Demple, B. (2000) 'Defining a rob Regulon in Escherichia coli by Using Transposon Mutagenesis', *Journal of Bacteriology*, 182(13), pp. 3794–3801. doi: 10.1128/JB.182.13.3794-3801.2000.Updated.
- Bennish, M. L. (1991) 'Potentially lethal complications of shigellosis.', *Reviews of infectious diseases*, 13 Suppl 4, pp. S319–S324.
- Berry, J.-L. and Pelicic, V. (2015) 'Exceptionally widespread nanomachines composed of type IV pilins: the prokaryotic Swiss Army knives', *FEMS Microbiology Reviews*, 39(1), pp. 1–21. doi: 10.1093/femsre/fuu001.
- Bhattacharya, S., Sarkar, K., Nair, G. B., Faruque, A. and Sack, D. A. (2003) 'Multidrug-resistant Shigella dysenteriae type 1 in south Asia', *Lancet Infectious Diseases*, 3(December), p. 2003.

- Bleibtreu, A., Clermont, O., Darlu, P., Glodt, J., Branger, C., Picard, B. and Denamur, E. (2014) 'The rpoS gene is predominantly inactivated during laboratory storage and undergoes source-sink evolution in Escherichia coli species', *Journal of Bacteriology*, 196(24), pp. 4276–4284. doi: 10.1128/JB.01972-14.
- Bodhidatta, L., Vithayasai, N., Eimpokalarp, B., Pitarangsi, C., Serichantalergs, O. and Isenbarger, D. W. (2002) 'Bacterial enteric pathogens in children with acute dysentery in thailand: Increasing importance of quinolone-resistant Campylobacter', *Southeast Asian Journal of Tropical Medicine and Public Health*, 33(4), pp. 752–757.
- Van Boeckel, T. P., Gandra, S., Ashok, A., Caudron, Q., Grenfell, B. T., Levin, S. a. and Laxminarayan, R. (2014) 'Global antibiotic consumption 2000 to 2010: An analysis of national pharmaceutical sales data', *The Lancet Infectious Diseases*. Elsevier Ltd, 14(8), pp. 742–750. doi: 10.1016/S1473-3099(14)70780-7.
- Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. and Pirovano, W. (2011) 'Scaffolding pre-assembled contigs using SSPACE', *Bioinformatics*, 27(4), pp. 578–579. doi: 10.1093/bioinformatics/btq683.
- Boetzer, M. and Pirovano, W. (2012) 'Toward almost closed genomes with GapFiller', *Genome Biology*. BioMed Central Ltd, 13(6), p. R56. doi: 10.1186/gb-2012-13-6-r56.
- Bollback, J. P. (2006) 'SIMMAP: stochastic character mapping of discrete traits on phylogenies.', *BMC bioinformatics*, 7, p. 88. doi: 10.1186/1471-2105-7-88.
- Borg, M. L., Modi, a, Tostmann, a, Gobin, M., Cartwright, J., Quigley, C., Crook, P., Boxall, N., Paul, J., Cheasty, T., Gill, N., Hughes, G., Simms, I. and Oliver, I. (2012) 'Ongoing outbreak of Shigella flexneri serotype 3a in men who have sex with men in England and Wales, data from 2009-2011.', *Euro surveillance : bulletin Européen sur les maladies transmissibles = European communicable disease bulletin*, 17(13), pp. 1–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22490381.
- Bose, R., Nashipuri, J. N., Sen, P. K., Datta, P., Bhattacharya, S. K., Datta, D., Sen, D. and Bhattacharya, M. K. (1984) 'Epidemic of Dysentery in West Bengal: Clinicians' Enigma', *The Lancet*, p. 1160. doi: 10.1016/S0140-6736(84)91600-3.

- Bowen, A., Hurd, J., Hoover, C., Khachadourian, Y., Traphagen, E., Harvey, E., Libby, T., Ehlers, S., Ongpin, M., Norton, J., Bicknese, A. and Kimura, A. (2015) 'Importation and Domestic Transmission of Shigella sonnei Resistant to Ciprofloxacin - United States, May 2014-February 2015', *Morbidity and Mortality Weekly Report*, 64(12), pp. 318–320.
- Bradshaw, D. J., Marsh, P. D., Keith Watson, G. and Allison, C. (1998) 'Role of Fusobacterium nucleatum and coaggregation in anaerobe survival in planktonic and biofilm oral microbial communities during aeration', *Infection and Immunity*, 66(10), pp. 4729–4732.
- Bravo, V., Puhar, A., Sansonetti, P., Parsot, C. and Toro, C. S. (2015) 'Distinct Mutations Led to Inactivation of Type 1 Fimbriae Expression in Shigella spp.', *Plos One*, 10(3), p. e0121785. doi: 10.1371/journal.pone.0121785.
- Breiman, L. (2001) 'Random forests', *Machine learning*, pp. 5–32. doi: 10.1023/A:1010933404324.
- Brotcke Zumsteg, A., Goosmann, C., Brinkmann, V., Morona, R. and Zychlinsky, A. (2014) 'IcsA is a Shigella flexneri adhesin regulated by the type III secretion system and required for pathogenesis.', *Cell host & microbe*. Elsevier, 15(4), pp. 435–45. doi: 10.1016/j.chom.2014.03.001.
- Browne, H. P., Neville, B. A., Forster, S. C. and Lawley, T. D. (2017) 'Transmission of the gut microbiota: spreading of health', *Nature Reviews Microbiology*. doi: 10.1038/nrmicro.2017.50.
- Buchrieser, C., Glaser, P., Rusniok, C., Nedjari, H., D'Hauteville, H., Kunst, F., Sansonetti, P. and Parsot, C. (2000) 'The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of Shigella flexneri', *Molecular microbiology*. 2000/12/15, 38(4), pp. 760–771. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11115111.
- Buffie, C. G., Bucci, V., Stein, R. R., McKenney, P. T., Ling, L., Gobourne, A., No, D., Liu, H., Kinnebrew, M., Viale, A., Littmann, E., van den Brink, M. R. M., Jenq, R. R., Taur, Y., Sander, C., Cross, J., Toussaint, N. C., Xavier, J. B. and Pamer, E. G. (2014) 'Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile', *Nature*. Nature Publishing Group, 517(7533), pp. 205–8. doi:

10.1038/nature13828.

- Calcuttawala, F., Hariharan, C., Pazhani, G. P., Ghosh, S. and Ramamurthy, T. (2015) 'Activity Spectrum of Colicins Produced by Shigella sonnei and Genetic Mechanism of Colicin Resistance in Conspecific S. sonnei Strains and Escherichia coli', *Antimicrobial Agents and Chemotherapy*, 59(1), pp. 152–158. doi: 10.1128/AAC.04122-14.
- Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L. and Darst, S. A. (2002) 'Structure of the bacterial RNA polymerase promoter specificity σ subunit', *Molecular Cell*. Cell Press, 9(3), pp. 527–539. doi: 10.1016/S1097-2765(02)00470-7.
- Carattoli, A., Zankari, E., Garciá-Fernández, A., Larsen, M. V., Lund, O., Villa, L., Aarestrup, F. M. and Hasman, H. (2014) 'In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing', *Antimicrobial Agents and Chemotherapy*, 58(7), pp. 3895–3903. doi: 10.1128/AAC.02412-14.
- Carayol, N. and Tran Van Nhieu, G. (2013) 'Tips and tricks about Shigella invasion of epithelial cells.', *Current opinion in microbiology*. Elsevier Ltd, 16(1), pp. 32–7. doi: 10.1016/j.mib.2012.11.010.
- Carvalho, F. a, Aitken, J. D., Gewirtz, a T. and Vijay-Kumar, M. (2011) 'TLR5 activation induces secretory interleukin-1 receptor antagonist (sIL-1Ra) and reduces inflammasome-associated tissue damage.', *Mucosal immunology*. Nature Publishing Group, 4(1), pp. 102–111. doi: 10.1038/mi.2010.57.
- Carver, T., Berriman, M., Tivey, A., Patel, C., Böhme, U., Barrell, B. G., Parkhill, J. and Rajandream, M. A. (2008) 'Artemis and ACT: Viewing, annotating and comparing sequences stored in a relational database', *Bioinformatics*, 24(23), pp. 2672–2676. doi: 10.1093/bioinformatics/btn529.
- Charbonneau, M. R., Blanton, L. V, DiGiulio, D. B., Relman, D. A., Lebrilla, C. B., Mills, D. A. and Gordon, J. I. (2016) 'A microbial perspective of human developmental biology', *Nature*, 535(7610), pp. 48–55. doi: 10.1038/nature18845.
- Chattaway, M. A., Greig, D. R., Gentle, A., Hartman, H. B., Dallman, T. J. and Jenkins, C. (2017) 'Whole-genome sequencing for national surveillance of Shigella flexneri', *Frontiers in Microbiology*, 8(SEP), pp. 1–6. doi:

10.3389/fmicb.2017.01700.

- Checkley, W., Buckley, G., Gilman, R. H., Assis, A. M., Guerrant, R. L., Morris, S. S., Molbak, K., Valentiner-Branth, P., Lanata, C. F. and Black, R. E. (2008) 'Multi-country analysis of the effects of diarrhoea on childhood stunting', *International Journal of Epidemiology*, 37(4), pp. 816–830. doi: 10.1093/ije/dyn099.
- Checroun, C., Bordes, P., Leroy, O., Kolb, A. and Gutierrez, C. (2004) 'Interactions between the 2.4 and 4.2 regions of σS, the stress-specific σ factor of Escherichia coli, and the -10 and -35 promoter elements', *Nucleic Acids Research*, 32(1), pp. 45–53. doi: 10.1093/nar/gkh155.
- Cheng, L., Connor, T. R., Sirén, J., Aanensen, D. M. and Corander, J. (2013) 'Hierarchical and spatially explicit clustering of DNA sequences with BAPS software', *Molecular Biology and Evolution*, pp. 1224–1228. doi: 10.1093/molbev/mst028.
- Chewapreecha, C., Holden, M. T. G., Vehkala, M., Välimäki, N., Yang, Z., Harris, S. R., Mather, A. E., Tuanyok, A., De Smet, B., Le Hello, S., Bizet, C., Mayo, M., Wuthiekanun, V., Limmathurotsakul, D., Phetsouvanh, R., Spratt, B. G., Corander, J., Keim, P., Dougan, G., Dance, D. A. B., Currie, B. J., Parkhill, J., Peacock, S. J., Limmathurotsakul, D., Nandi, T., Johnson, S. L., Holden, M. T. G., Price, E. P., Pearson, T., Page, A. J., Dale, J., Gee, J. E., Allender, C. J., Tuanyok, A., Elrod, M. G., Hoffmaster, A. R., Sarovich, D. S., Croucher, N. J., Drummond, A. J., Suchard, M. A., Xie, D., Rambaut, A., Liu, J. H., Lawrence, B., Ward, C., Abraham, S., Currie, B. J., Lees, J. A., Sarovich, D. S., Tuanyok, A., French, C. T., Benanti, E. L., Nguyen, C. M., Welch, M. D., Tuanyok, A., Willcocks, S. J., Denman, C. C., Atkins, H. S., Wren, B. W., Chen, Y., Bast, A., Wallace, A. R., Wood, D. E., Salzberg, S. L., Zerbino, D. R., Birney, E., Chewapreecha, C., Seemann, T., Nandi, T., Spring-Pearson, S. M., Price, M. N., Dehal, P. S., Arkin, A. P., Harris, S. R., Page, A. J., Stamatakis, A., Corander, J., Marttinen, P., Sirén, J., Tang, J., Cheng, L., Connor, T. R., Siren, J., Aanensen, D. M., Corander, J., Croucher, N. J., Assefa, S., Keane, T. M., Otto, T. D., Newbold, C., Berriman, M., Carver, T. J., Harris, S. R., Ho, S. Y. W., Duchene, S., Murray, G. G. R., Lieberman, T. D., Mathers, A. J., Young, B. C., Croucher, N. J., Revell, L. J., Bollback, J. P., Laabei, M., Chewapreecha, C., Kent,

W. J., Ooi, W. F., Winsor, G. L., Letunic, I. and Bork, P. (2017) 'Global and regional dissemination and evolution of Burkholderia pseudomallei', *Nature Microbiology*, 2(January), p. 16263. doi: 10.1038/nmicrobiol.2016.263.

- Chiou, C.-S., Izumiya, H., Kawamura, M., Liao, Y.-S., Su, Y.-S., Wu, H.-H., Chen, W.-C. and Lo, Y.-C. (2016) 'The worldwide spread of ciprofloxacin-resistant Shigella sonnei among HIV-infected men who have sex with men, Taiwan', *Clinical Microbiology and Infection*, 22(4), p. 383.e11-383.e16. doi: 10.1016/j.cmi.2015.12.021.
- Choi, S. Y., Jeon, Y. S., Lee, J. H., Choi, B., Moon, S. H., von Seidlein, L., Clemens, J. D., Dougan, G., Wain, J., Yu, J., Lee, J. C., Seol, S. Y., Lee, B. K., Song, J. H., Song, M., Czerkinsky, C., Chun, J. and Kim, D. W. (2007) 'Multilocus sequence typing analysis of Shigella flexneri isolates collected in Asian countries', *Journal of medical microbiology*. 2007/10/30, 56(Pt 11), pp. 1460–1466. doi: 10.1099/jmm.0.47322-0.
- Christopher, P., Kv, D., Sm, J. and Sankarapandian, V. (2010) 'Antibiotic therapy for Shigella dysentery (Review)', *Cochrane Database Syst Rev*, (8), p. Cd006784. doi: 10.1002/14651858.CD006784.pub3\r10.1002/14651858.CD006784.pub3.
- Chubiz, L. M., Glekas, G. D. and Rao, C. V. (2012) 'Transcriptional cross talk within the Mar-Sox-Rob regulon in Escherichia coli is limited to the rob and marRAB operons', *Journal of Bacteriology*, 194(18), pp. 4867–4875. doi: 10.1128/JB.00680-12.
- Chung The, H., Pham Thanh, D., Holt, K. E., Thomson, N. R. and Baker, S. (2016) 'The genomic signatures of Shigella evolution, adaptation and geographical spread', *Nature Reviews Microbiology*. Nature Publishing Group, 14(4), pp. 235–250. doi: 10.1038/nrmicro.2016.10.
- Chung The, H., Rabaa, M., Pham Thanh, D., Ruekit, S., Wangchuk, S., Dorji, T., Tshering, K. P., Nguyen, T. N. T., Vinh, P., Thanh, T., Minh, C., Turner, P., Sar, P., Thwaites, G., Holt, K. E., Thomson, N. R., Bodhidatta, L., Mason, C. J. and Baker, S. (2015) 'Introduction and establishment of fluoroquinolone resistant Shigella sonnei into Bhutan', *Microbial Genomics*, 1(6), pp. 1–11. doi: 10.1099/mgen.0.000042.

- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X. and Ruden, D. M. (2012) 'A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff', *Fly*, 6(2), pp. 80–92. doi: 10.4161/fly.19695.
- Clark, C., Beltrame, J. and Manning, P. (1991) 'The oac gene encoding a lipopolysaccharide O-antigen acetylase maps adjacent to the integrase-encoding gene on the genome of Shigella flexneri bacteriophage Sf6', *Gene*, 107, pp. 43–52. Available at: http://www.sciencedirect.com/science/article/pii/037811199190295M (Accessed: 5 December 2014).
- CLSI (2012) *Clinical and Laboratory Standards Institute: Performance standards for antimicrobial susceptibility testing; twentieth informational supplement.*
- Connor, T. R., Barker, C. R., Baker, K. S., Weill, F.-X., Talukder, K. A., Smith, A. M., Baker, S., Gouali, M., Pham Thanh, D., Jahan Azmi, I., Dias da Silveira, W., Semmler, T., Wieler, L. H., Jenkins, C., Cravioto, A., Faruque, S. M., Parkhill, J., Wook Kim, D., Keddy, K. H. and Thomson, N. R. (2015) 'Species-wide whole genome sequencing reveals historical global spread and recent local persistence in *Shigella flexneri*', *eLife*, 4, pp. 1–16. doi: 10.7554/eLife.07335.
- Corander, J., Marttinen, P., Sirén, J. and Tang, J. (2008) 'Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations.', *BMC bioinformatics*, 9, p. 539. doi: 10.1186/1471-2105-9-539.
- Croucher, N. J., Page, a. J., Connor, T. R., Delaney, a. J., Keane, J. a., Bentley, S. D., Parkhill, J. and Harris, S. R. (2014) 'Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins', *Nucleic Acids Research*, 44(0), pp. 1–13. doi: 10.1093/nar/gku1196.
- Das, A., Natarajan, M. and Mandal, J. (2016) 'The Emergence of Quinolone Resistant Shigella sonnei, Pondicherry, India', *Plos One*, 11(8), p. e0160290. doi: 10.1371/journal.pone.0160290.
- David, L. A., Materna, A. C., Friedman, J., Campos-Baptista, M. I., Blackburn, M. C., Perrotta, A., Erdman, S. E. and Alm, E. J. (2014) 'Host lifestyle affects human microbiota on daily timescales.', *Genome biology*, 15(7), p. R89. doi: 10.1186/gb-2014-15-7-r89.

- David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V, Devlin, A. S., Varma, Y., Fischbach, M. A., Biddinger, S. B., Dutton, R. J. and Turnbaugh, P. J. (2014) 'Diet rapidly and reproducibly alters the human gut microbiome.', *Nature*. Nature Publishing Group, 505(7484), pp. 559–63. doi: 10.1038/nature12820.
- David, L., Weil, A., Ryan, E. T., Calderwood, S. B., Harris, J. B., Chowdhury, F., Begum, Y., Qadri, F., Larocque, R. C. and Turnbaugh, J. (2015) 'Gut Microbial Succession Follows Acute Secretory Diarrhea in Humans', *mBio*, 6(3), pp. 1–14. doi: 10.1128/mBio.00381-15.Editor.
- Deboer, M. D., Lima, A. A. M., Oría, R. B., Scharf, R. J., Moore, S. R., Luna, M. A. and Guerrant, R. L. (2012) 'Early childhood growth failure and the developmental origins of adult disease: Do enteric infections and malnutrition increase risk for the metabolic syndrome?', *Nutrition Reviews*, 70(11), pp. 642–653. doi: 10.1111/j.1753-4887.2012.00543.x.
- DiGiulio, D. B., Callahan, B. J., McMurdie, P. J., Costello, E. K., Lyell, D. J., Robaczewska, A., Sun, C. L., Goltsman, D. S. a., Wong, R. J., Shaw, G., Stevenson, D. K., Holmes, S. P. and Relman, D. a. (2015) 'Temporal and spatial variation of the human microbiota during pregnancy', *Proceedings of the National Academy of Sciences*, 112(35), p. 201502875. doi: 10.1073/pnas.1502875112.
- Dodd, C. E. and Jones, D. (1982) 'A numerical taxonomic study of the genus Shigella', *Journal of general microbiology*. 1982/09/01, 128(9), pp. 1933–1957. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7175494.
- Dogra, S., Sakwinska, O., Soh, S.-E., Ngom-bru, C., Brück, M., Berger, B., Brüssow, H., Karnani, N., Lee, Y. S., Yap, F., Chong, Y., Godfrey, K. M., Holbrook, J. D., Dogra, S., Sakwinska, O., Soh, S., Ngom-bru, C., Brück, M., Berger, B., Brüssow, H., Karnani, N., Lee, Y. S., Yap, F., Chong, Y., Godfrey, K. M., Holbrook, J. D., Dogra, S., Sakwinska, O., Soh, S., Ngom-bru, C., Brück, W. M., Lee, Y. S., Yap, F., Chong, S., Godfrey, K. M. and Joanna, D. (2015) 'Rate of establishing the gut microbiota in infancy has Consequences for Future Health', *Gut microbes*, 6(5), pp. 321–325. doi: 10.1080/19490976.2015.1078051.
- Dong, N., Zhu, Y., Lu, Q., Hu, L., Zheng, Y. and Shao, F. (2012) 'Structurally

distinct bacterial TBC-like GAPs link Arf GTPase to Rab1 inactivation to counteract host defenses.', *Cell*. Elsevier Inc., 150(5), pp. 1029–41. doi: 10.1016/j.cell.2012.06.050.

- Drekonja, D., Reich, J., Gezahegn, S., Greer, N., Shaukat, A., MacDonald, R., Rutks, I. and Wilt, T. J. (2015) 'Fecal microbiota transplantation for clostridium difficile infection a systematic review', *Annals of Internal Medicine*, 162(9), pp. 630–638. doi: 10.7326/M14-2693.
- Drummond, A. J., Ho, S. Y. W., Phillips, M. J. and Rambaut, A. (2006) 'Relaxed phylogenetics and dating with confidence.', *PLoS biology*, 4(5), p. e88. doi: 10.1371/journal.pbio.0040088.
- Drummond, A. J. and Rambaut, A. (2007) 'BEAST: Bayesian evolutionary analysis by sampling trees.', *BMC evolutionary biology*, 7, p. 214. doi: 10.1186/1471-2148-7-214.
- Drummond, A. J. and Suchard, M. a (2010) 'Bayesian random local clocks, or one rate to rule them all.', *BMC biology*, 8, p. 114. doi: 10.1186/1741-7007-8-114.
- Drummond, a J., Rambaut, a, Shapiro, B. and Pybus, O. G. (2005) 'Bayesian coalescent inference of past population dynamics from molecular sequences.', *Molecular biology and evolution*, 22(5), pp. 1185–92. doi: 10.1093/molbev/msi103.
- Dudley, E. G., Abe, C., Ghigo, J. M., Latour-Lambert, P., Hormazabal, J. C. and Nataro, J. P. (2006) 'An IncI1 plasmid contributes to the adherence of the atypical enteroaggregative Escherichia coli strain C1096 to cultured cells and abiotic surfaces', *Infection and Immunity*, 74(4), pp. 2102–2114. doi: 10.1128/IAI.74.4.2102-2114.2006.
- Duong, V. T., Phat, V. V., Tuyen, H. T., Dung, T. T. N., Trung, P. D., Minh, P. Van, Phuong Tu, L. T., Campbell, J. I., Phuc, H. Le, Thanh Ha, T. T., Ngoc, N. M., Huong, N. T. T., Tam, P. T. T., Huong, D. T., Xang, N. Van, Dong, N., Phuong, L. T., Hung, N. Van, Phu, B. D., Phuc, T. M., Thwaites, G. E., Vi, L. L., Rabaa, M. A., Thompson, C. N. and Baker, S. (2016) 'Evaluation of luminex xTAG gastrointestinal pathogen panel assay for detection of multiple diarrheal pathogens in fecal samples in Vietnam', *Journal of Clinical Microbiology*, 54(4), pp. 1094–1100. doi: 10.1128/JCM.03321-15.

- Duong, V. T., Tuyen, H. T., Minh, P. Van, Campbell, J. I., Phuc, H. Le, Do, T., Nhu, H., Thi, L., Tu, P., Thi, T. and Chau, H. (2018) 'No Clinical Benefit of Empirical Antimicrobial Therapy for Pediatric Diarrhea in a High-Usage , High-Resistance Setting', *Clinical Infectious Diseases*, 66(March). doi: 10.1093/cid/cix844.
- Duval V. and Lister, I. (2014) 'MarA, SoxS and Rob of Escherichia coli – Global regulators of multidrug resistance, virulence and stress response', *Int J Biotechnol Wellness Ind.*, 2(3), pp. 101–124. doi: 10.6000/1927-3037.2013.02.03.2.MarA.
- Edgar, R. C. (2013) 'UPARSE: highly accurate OTU sequences from microbial amplicon reads.', *Nature methods*, 10(10), pp. 996–8. doi: 10.1038/nmeth.2604.
- Eftekhari, N., Bakhshi, B., Pourshafie, M. R., Zarbakhsh, B., Rahbar, M., Hajia, M. and Ghazvini, K. (2013) 'Genetic Diversity of *Shigella* spp. and Their Integron Content', *Foodborne Pathogens and Disease*, 10(3), pp. 237–242. doi: 10.1089/fpd.2012.1250.
- Egendre, P. I. L. and Anderson, M. J. (1999) 'Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments', *Ecological Monographs*, 69(1), pp. 1–24.
- Egile, C., Loisel, T. and Laurent, V. (1999) 'Activation of the CDC42 effector N-WASP by the Shigella flexneri IcsA protein promotes actin nucleation by Arp2/3 complex and bacterial actin-based motility', *The Journal of Cell Biology*, 146(6), pp. 1319–1332. Available at: http://jcb.rupress.org/content/146/6/1319.abstract (Accessed: 9 January 2015).
- Eiglmeier, K., Parkhill, J., Honoré, N., Garnier, T., Tekaia, F., Telenti, a, Klatser, P., James, K. D., Thomson, N. R., Wheeler, P. R., Churcher, C., Harris, D., Mungall, K., Barrell, B. G. and Cole, S. T. (2001) 'The decaying genome of Mycobacterium leprae.', *Leprosy review*, 72(4), pp. 387–398.
- El-Gendy, a. M., Mansour, a., Weiner, M. a., Pimentel, G., Armstrong, a. W., Young, S. Y. N., Elsayed, N. and Klena, J. D. (2012) 'Genetic diversity and antibiotic resistance in Shigella dysenteriae and Shigella boydii strains isolated from children aged <5 years in Egypt', *Epidemiology and Infection*, 140(2), pp. 299–310. doi: 10.1017/S0950268811000525.
- Fahad, B. M., Matin, A., Shill, M. C. and Asish, K. D. (2010) 'Antibiotic usage at a

primary health care unit in Bangladesh', *Australasian Medical Journal*, 3(7), pp. 414–421. doi: 10.4066/AMJ.2010.322.

- Faherty, C. S., Faherty, C., Harper, J. M., Shea-Donohue, T., Barry, E. M., Kaper, J. B., Fasano, A. and Nataro, J. P. (2012) 'Chromosomal and plasmid-encoded factors of Shigella flexneri induce secretogenic activity ex vivo.', *PloS one*, 7(11), p. e49980. doi: 10.1371/journal.pone.0049980.
- Falkow, S. (1958) 'Activity of lysine decarboxylase as an aid in the identification of salmonellae and shigellae', *Am. J. Clin. Pathol.*, 29, pp. 598–600.
- Fan, Y., Wu, J., Ung, M. H., Lay, N. De, Cheng, C. and Ling, J. (2015) 'Protein mistranslation protects bacteria against oxidative stress', *Nucleic acids research*, 43(3), pp. 1740–1748. doi: 10.1093/nar/gku1404.
- Fasano, a, Noriega, F. R., Liao, F. M., Wang, W. and Levine, M. M. (1997) 'Effect of shigella enterotoxin 1 (ShET1) on rabbit intestine in vitro and in vivo.', *Gut*, 40(4), pp. 505–511. doi: 10.1136/gut.40.4.505.
- Feng, Y., Chen, Z. and Liu, S.-L. (2011) 'Gene decay in Shigella as an incipient stage of host-adaptation.', *PloS one*, 6(11), p. e27754. doi: 10.1371/journal.pone.0027754.
- Fernandez-prada, C. M., Hoover, D. L., Tall, B. E. N. D., Hartman, A. B., Kopelowitz, J. and Venkatesan, M. M. (2000) 'Shigella flexneri IpaH7.8 Facilitates Escape of Virulent Bacteria from the Endocytic Vacuoles of Mouse and Human Macrophages', *Infection and immunity*, 68(6), pp. 3608–3619.
- Ferreccio, C., Prado, V., Ojeda, A., Cayyazo, M., Abrego, P., Guers, L. and Levine, M. M. (1991) 'Epidemiologic Patterns of Acute Diarrhea and Endemic Shigella Infections in Children in a Poor Periurban Setting in', *American Journal of Epidemiology*, 134(6), pp. 614–627.
- Fisher, C. R., Davies, N. M. L. L., Wyckoff, E. E., Feng, Z., Oaks, E. V and Payne, S. M. (2009) 'Genetics and virulence association of the Shigella flexneri sit iron transport system.', *Infection and immunity*, 77(5), pp. 1992–9. doi: 10.1128/IAI.00064-09.
- Flexner, S. (1900) 'On the etiology of tropical dysentery', *Philadelphia Medical Journal*, 6, pp. 414–421.

- Fontaine, F., Stewart, E. J., Lindner, A. B. and Taddei, F. (2008) 'Mutations in two global regulators lower individual mortality in Escherichia coli', *Molecular Microbiology*, 67(1), pp. 2–14. doi: 10.1111/j.1365-2958.2007.05988.x.
- Friedman, J. and Alm, E. J. (2012) 'Inferring correlation networks from genomic survey data.', *PLoS computational biology*, 8(9), p. e1002687. doi: 10.1371/journal.pcbi.1002687.
- Fukuyama, J., Rumker, L., Sankaran, K., Jeganathan, P., Dethlefsen, L., Relman, D. A. and Holmes, S. P. (2017) 'Multidomain analyses of a longitudinal human microbiome intestinal cleanout perturbation experiment', *PLoS computational biology*, pp. 1–29.
- Fuzi, M., Szabo, D. and Csercsik, R. (2017) 'Double-serine fluoroquinolone resistance mutations advance major international clones and lineages of various multi-drug resistant bacteria', *Frontiers in Microbiology*, 8(NOV), pp. 1–14. doi: 10.3389/fmicb.2017.02261.
- Gaudreau, C., Ratnayake, R., Pilon, P. a, Gagnon, S., Roger, M. and Lévesque, S. (2011) 'CiprofloxacinResistant Shigella sonnei among Men Who Have Sex with Men, Canada, 2010', *Emerging Infectious Diseases*, 17(9), pp. 1747–1750.
- Gentle, A., Ashton, P. M., Dallman, T. J. and Jenkins, C. (2016) 'Evaluation of Molecular Methods for Serotyping Shigella flexneri', *Journal of Clinical Microbiology*, 54(6), pp. 1456–1461. doi: 10.1128/JCM.03386-15.Editor.
- Ghosh, S., Pazhani, G. P., Chowdhury, G., Guin, S., Dutta, S., Rajendran, K., Bhattacharya, M. K., Takeda, Y., Niyogi, S. K., Balakrish Nair, G. and Ramamurthy, T. (2011) 'Genetic characteristics and changing antimicrobial resistance among Shigella spp. isolated from hospitalized diarrhoeal patients in Kolkata, India', *Journal of Medical Microbiology*, 60(10), pp. 1460–1466. doi: 10.1099/jmm.0.032920-0.
- Gilbert, J. A., Quinn, R. A., Debelius, J., Xu, Z. Z., Morton, J., Garg, N., Jansson, J. K., Dorrestein, P. C. and Knight, R. (2016) 'Microbiome-wide association studies link dynamic microbial consortia to disease', *Nature*, 535(7610), pp. 94–103. doi: 10.1038/nature18850.
- Girón, J. a (1995) 'Expression of flagella and motility by Shigella.', *Molecular*

microbiology, 18(1), pp. 63–75. doi: 10.1111/j.1365-2958.1995.mmi_18010063.x.

- Grad, Y. H., Harris, S. R., Kirkcaldy, R. D., Green, A. G., Marks, D. S., Bentley, S. D., Trees, D. and Lipsitch, M. (2016) 'Genomic epidemiology of gonococcal resistance to extended-spectrum cephalosporins, macrolides, and fluoroquinolones in the United States, 2000-2013.', *J Infect Dis*, 214(10), pp. 1579–87. doi: 10.1093/infdis/jiw420.
- Gray, M. D., Lacher, D. W., Leonard, S. R., Abbott, J., Zhao, S., Lampel, K. A., Prothery, E., Gouali, M., Weill, F.-X. and Maurelli, A. T. (2015) 'Prevalence of Shiga toxin-producing Shigella species isolated from French travellers returning from the Caribbean: an emerging pathogen with international implications', *Clinical Microbiology and Infection*, 21(8), p. 765.e9-765.e14. doi: 10.1016/j.cmi.2015.05.006.
- Gray, M. D., Lampel, K. A., Strockbine, N. A., Fernandez, R. E., Melton-celsa, A. R. and Maurelli, A. T. (2014) 'Clinical Isolates of Shiga Toxin 1a – Producing Shigella flexneri with an Epidemiological Link to Recent Travel to Hispañiola', *Emerging infectious diseases*, 20(10), pp. 1669–1677.
- Gray, M., Leonard, S., Lacher, D., Lampel, K., Alam, M., Morris Jr., J., Ali, A., LaBreck, P. and Maurelli, A. T. (2015) 'Stx-Producing Shigella Species From Patients in Haiti: An Emerging Pathogen With the Potential for Global Spread', *Open Forum Infectious Diseases*, pp. 1–8. doi: 10.1093/o.
- Griffith, K. L., Fitzpatrick, M. M., Keen, E. F. and Wolf, R. E. (2009) 'Two Functions of the C-Terminal Domain of Escherichia coli Rob: Mediating "Sequestration-Dispersal" as a Novel Off-On Switch for Regulating Rob's Activity as a Transcription Activator and Preventing Degradation of Rob by Lon Protease', *Journal of Molecular Biology*. Elsevier Ltd, 388(3), pp. 415–430. doi: 10.1016/j.jmb.2009.03.023.
- Gu, B., Cao, Y., Pan, S., Zhuang, L., Yu, R., Peng, Z., Qian, H., Wei, Y., Zhao, L., Liu, G. and Tong, M. (2012) 'Comparison of the prevalence and changing resistance to nalidixic acid and ciprofloxacin of Shigella between Europe-America and Asia-Africa from 1998 to 2009', *International journal of antimicrobial agents*. 2012/04/10, 40(1), pp. 9–17. doi: 10.1016/j.ijantimicag.2012.02.005.

- Guan, S., Bastin, D. and Verma, N. (1999) 'Functional analysis of the O antigen glucosylation gene cluster of Shigella flexneri bacteriophage SfX', *Microbiology*, 145, pp. 1263–1273. Available at: http://mic.sgmjournals.org/content/145/5/1263.short (Accessed: 5 December 2014).
- Guerin, P. J., Brasher, C., Baron, E., Mic, D., Grimont, F., Ryan, M., Aavitsland, P. and Legros, D. (2003) 'Shigella dysenteriae serotype 1 in west Africa : intervention strategy for an outbreak in Sierra Leone', *the Lancet*, 362, pp. 705–706.
- Guerrant, R. L., DeBoer, M. D., Moore, S. R., Scharf, R. J. and Lima, A. A. M. (2013) 'The impoverished gut--a triple burden of diarrhoea, stunting and chronic disease.', *Nature reviews. Gastroenterology & hepatology*, 10(4), pp. 220–9. doi: 10.1038/nrgastro.2012.239.
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010) 'New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.', *Systematic biology*, 59(3), pp. 307–21. doi: 10.1093/sysbio/syq010.
- Gupta, R. S. and Bhandari, V. (2014) 'Phylogeny and molecular signatures for the phylum Fusobacteria and its distinct subclades', *Anaerobe*. Elsevier Ltd, 28(1), pp. 182–198. doi: 10.1007/s10482-011-9576-z.
- Gupta, S. K., Strockbine, N., Omondi, M., Hise, K., Fair, M. A. and Mintz, E. (2007) 'Short report: Emergence of Shiga toxin 1 genes within Shigella dysenteriae type 4 isolates from travelers returning from the Island of Hispa??ola', *American Journal of Tropical Medicine and Hygiene*, 76(6), pp. 1163–1165.
- Halfvarson, J., Brislawn, C. J., Lamendella, R., Vázquez-Baeza, Y., Walters, W. A., Bramer, L. M., D'Amato, M., Bonfiglio, F., McDonald, D., Gonzalez, A., McClure, E. E., Dunklebarger, M. F., Knight, R. and Jansson, J. K. (2017) 'Dynamics of the human gut microbiome in inflammatory bowel disease', *Nature Microbiology*, 2(February), p. 17004. doi: 10.1038/nmicrobiol.2017.4.
- Han, J., Lynne, A. M., David, D. E., Tang, H., Xu, J., Nayak, R., Kaldhone, P., Logue, C. M. and Foley, S. L. (2012) 'DNA Sequence Analysis of Plasmids from Multidrug Resistant Salmonella enterica Serotype Heidelberg Isolates', *PLoS ONE*, 7(12), pp. 1–8. doi: 10.1371/journal.pone.0051160.

- Handa, Y., Suzuki, M., Ohya, K., Iwai, H., Ishijima, N., Koleske, A. J., Fukui, Y. and Sasakawa, C. (2007) 'Shigella IpgB1 promotes bacterial entry through the ELMO-Dock180 machinery.', *Nature cell biology*, 9(1), pp. 121–8. doi: 10.1038/ncb1526.
- Harrington, A., Darboe, N. and Kenjale, R. (2006) 'Characterization of the interaction of single tryptophan containing mutants of IpaC from Shigella flexneri with phospholipid membranes', *Biochemistry*, 45, pp. 626–636. Available at: http://pubs.acs.org/doi/abs/10.1021/bi0512593 (Accessed: 9 January 2015).
- He, F., Han, K., Liu, L., Sun, W., Zhang, L., Zhu, B. and Ma, H. (2012) 'Shigellosis Outbreak Associated with Contaminated Well Water in a Rural Elementary School: Sichuan Province, China, June 7-16, 2009', *PLoS ONE*, 7(10), pp. 16–19. doi: 10.1371/journal.pone.0047239.
- He, M., Miyajima, F., Roberts, P., Ellison, L., Pickard, D. J., Martin, M. J., Connor, T. R., Harris, S. R., Fairley, D., Bamford, K. B., D'Arc, S., Brazier, J., Brown, D., Coia, J. E., Douce, G., Gerding, D., Kim, H. J., Koh, T. H., Kato, H., Senoh, M., Louie, T., Michell, S., Butt, E., Peacock, S. J., Brown, N. M., Riley, T., Songer, G., Wilcox, M., Pirmohamed, M., Kuijper, E., Hawkey, P., Wren, B. W., Dougan, G., Parkhill, J. and Lawley, T. D. (2013) 'Emergence and global spread of epidemic healthcare-associated Clostridium difficile.', *Nature genetics*. Nature Publishing Group, 45(1), pp. 109–13. doi: 10.1038/ng.2478.
- Hedge, J. and Wilson, D. J. (2016) 'Practical Approaches for Detecting Selection in Microbial Genomes', *PLOS Computational Biology*, 12(2), p. e1004739. doi: 10.1371/journal.pcbi.1004739.
- Hershberg, R., Tang, H. and Petrov, D. A. (2007) 'Reduced selection leads to accelerated gene loss in Shigella', *Genome biology*. 2007/08/10, 8(8), p. R164. doi: 10.1186/gb-2007-8-8-r164.
- Hilbi, H., Moss, J. E., Hersh, D., Chen, Y., Arondel, J., Banerjee, S., Flavell, R. A., Yuan, J., Sansonetti, P. J. and Zychlinsky, A. (1998) 'Shigella-induced Apoptosis Is Dependent on Caspase-1 Which Binds to IpaB *', *The Journal of biological chemistry*, 273(49), pp. 32895–32900.
- Holt, K. E., Baker, S., Weill, F. X., Holmes, E. C., Kitchen, A., Yu, J., Sangal, V.,

Brown, D. J., Coia, J. E., Kim, D. W., Choi, S. Y., Kim, S. H., da Silveira, W. D., Pickard, D. J., Farrar, J. J., Parkhill, J., Dougan, G. and Thomson, N. R. (2012) 'Shigella sonnei genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe', *Nature genetics*. 2012/08/07, 44(9), pp. 1056–1059. doi: 10.1038/ng.2369.

- Holt, K. E., Parkhill, J., Mazzoni, C. J., Roumagnac, P., Weill, F. X., Goodhead, I., Rance, R., Baker, S., Maskell, D. J., Wain, J., Dolecek, C., Achtman, M. and Dougan, G. (2008) 'High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi', *Nature genetics*. 2008/07/29, 40(8), pp. 987–993. doi: 10.1038/ng.195.
- Holt, K. E., Vu, T., Nga, T., Pham, D., Vinh, H., Wook, D., Phan, M. and Tra, V. (2013) 'Tracking the establishment of local endemic populations of an emergent enteric pathogen', *PNAS*, 110(43). doi: 10.1073/pnas.1308632110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1308632110.
- Hong, M. and Payne, S. M. (1997) 'Effect of mutations in Shigella flexneri chromosomal and plasmid-encoded lipopolysaccharide genes on invasion and serum resistance.', *Molecular microbiology*, 24(4), pp. 779–791.
- Horiuchi, S., Inagaki, Y., Yamamoto, N., Okamura, N., Imagawa, Y. and Nakaya, R. (1993) 'Reduced susceptibilities of Shigella sonnei strains isolated from patients with dysentery to fluoroquinolones', *Antimicrobial Agents and Chemotherapy*, 37(11), pp. 2486–2489. doi: 10.1128/AAC.37.11.2486.
- Hsiao, A., Ahmed, A. M. S., Subramanian, S., Griffin, N. W., Drewry, L. L., Petri, W. A., Haque, R., Ahmed, T. and Gordon, J. I. (2014) 'Members of the human gut microbiota involved in recovery from Vibrio cholerae infection', *Nature*. Nature Publishing Group, 515(7527), pp. 423–426. doi: 10.1038/nature13738.
- Huan, P. T., Bastin, D. a, Whittle, B. L., Lindberg, A. a and Verma, N. K. (1997) 'Molecular characterization of the genes involved in O-antigen modification, attachment, integration and excision in Shigella flexneri bacteriophage SfV', *Gene*, 195(2), pp. 217–227. doi: 10.1016/S0378-1119(97)00143-1.
- Hunt, M., Mather, A. E., Sánchez-Busó, L., Page, A. J., Parkhill, J., Keane, J. A. and Harris, S. R. (2017) 'ARIBA: rapid antimicrobial resistance genotyping directly from

sequencing reads', *bioRxiv*, p. 118000. doi: 10.1101/118000.

- Huseby, D. L., Pietsch, F., Brandis, G., Garoff, L., Tegehall, A. and Hughes, D. (2017) 'Mutation Supply and Relative Fitness Shape the Genotypes of Ciprofloxacin-Resistant Escherichia coli', *Molecular Biology and Evolution*, 34(5), pp. 1029–1039. doi: 10.1093/molbev/msx052.
- Hyma, K. E., Lacher, D. W., Nelson, A. M., Bumbaugh, A. C., Janda, J. M., Strockbine, N. A., Young, V. B. and Whittam, T. S. (2005) 'Evolutionary genetics of a new pathogenic Escherichia species: Escherichia albertii and related Shigella boydii strains', *Journal of bacteriology*. 2005/01/05, 187(2), pp. 619–628. doi: 10.1128/JB.187.2.619-628.2005.
- Iguchi, A., Iyoda, S., Seto, K. and Ohnishi, M. (2011) 'Emergence of a novel Shiga toxin-producing Escherichia coli O serogroup cross-reacting with Shigella boydii type 10', *Journal of Clinical Microbiology*, 49(10), pp. 3678–3680. doi: 10.1128/JCM.01197-11.
- Ingersoll, M., Groisman, E. A. and Zychlinsky, A. (2002) 'Pathogenicity islands of Shigella', *Current topics in microbiology and immunology*. 2002/05/17, 264(1), pp. 49–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12014181.
- Ishwaran, H. and Kogalur, U. B. (2016) 'Random Forests for Survival, Regression and Classification (RF-SRC), R package version 2.2.0'.
- Iwai, H., Kim, M., Yoshikawa, Y., Ashida, H., Ogawa, M., Fujita, Y., Muller, D., Kirikae, T., Jackson, P. K., Kotani, S. and Sasakawa, C. (2007) 'A bacterial effector targets Mad2L2, an APC inhibitor, to modulate host cell cycling.', *Cell*, 130(4), pp. 611–23. doi: 10.1016/j.cell.2007.06.043.
- Izumiya, H., Tada, Y., Ito, K., Morita-Ishihara, T., Ohnishi, M., Terajima, J. and Watanabe, H. (2009) 'Characterization of Shigella sonnei isolates from travel-associated cases in Japan', *Journal of Medical Microbiology*, 58(11), pp. 1486–1491. doi: 10.1099/jmm.0.011809-0.
- Jakhetia, R., Talukder, K. a and Verma, N. K. (2013) 'Isolation, characterization and comparative genomics of bacteriophage SfIV: a novel serotype converting phage from Shigella flexneri.', *BMC genomics*. BMC Genomics, 14(1), p. 677. doi: 10.1186/1471-2164-14-677.

- Jakobsson, H. E., Abrahamsson, T. R., Jenmalm, M. C., Harris, K., Quince, C., Jernberg, C., Björkstén, B., Engstrand, L. and Andersson, A. F. (2014) 'Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section.', *Gut*, 63(4), pp. 559–66. doi: 10.1136/gutjnl-2012-303249.
- Jandhyala, S. M., Talukdar, R., Subramanyam, C., Vuyyuru, H., Sasikala, M. and Reddy, D. N. (2015) 'Role of the normal gut microbiota', *World Journal of Gastroenterology*, 21(29), pp. 8836–8847. doi: 10.3748/wjg.v21.i29.8787.
- Jiang, Y., Yang, F., Zhang, X., Yang, J., Chen, L., Yan, Y., Nie, H., Xiong, Z., Wang, J., Dong, J., Xue, Y., Xu, X., Zhu, Y., Chen, S. and Jin, Q. (2005) 'The complete sequence and analysis of the large virulence plasmid pSS of Shigella sonnei', *Plasmid*. 2005/08/27, 54(2), pp. 149–159. doi: 10.1016/j.plasmid.2005.03.002.
- Joshi, N. and Fass, J. (2011) 'Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle.'
- Kaas, R. S., Friis, C., Ussery, D. W. and Aarestrup, F. M. (2012) 'Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse Escherichia coli genomes.', *BMC genomics*. BMC Genomics, 13(1), p. 577. doi: 10.1186/1471-2164-13-577.
- Kalluri, P., Cummings, K. C., Abbott, S., Malcolm, G. B., Hutcheson, K., Beall, a, Joyce, K., Polyak, C., Woodward, D., Caldeira, R., Rodgers, F., Mintz, E. D. and Strockbine, N. (2004) 'Epidemiological features of a newly described serotype of Shigella boydii.', *Epidemiology and infection*, 132(4), pp. 579–583. doi: 10.1017/S0950268804002377.
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. and Jermiin, L. S. (2017) 'ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates', *Nature Methods*. Nature Publishing Group, in press(May), pp. 1–5. doi: 10.1038/nmeth.4285.
- Kentner, D., Martano, G., Callon, M., Chiquet, P., Brodmann, M., Burton, O., Wahlander, A., Nanni, P., Delmotte, N., Grossmann, J., Limenitakis, J., Schlapbach,

R., Kiefer, P., Vorholt, J. a, Hiller, S. and Bumann, D. (2014) 'Shigella reroutes host cell central metabolism to obtain high-flux nutrient supply for vigorous intracellular growth.', *Proceedings of the National Academy of Sciences of the United States of America*, 111(27), pp. 9929–34. doi: 10.1073/pnas.1406694111.

- Kern, W. V., Oethinger, M., Jellen-Ritter, A. S. and Levy, S. B. (2000) 'Non-target gene mutations in the development of fluoroquinolone resistance in Escherichia coli', *Antimicrobial Agents and Chemotherapy*, 44(4), pp. 814–820. doi: 10.1128/AAC.44.4.814-820.2000.
- Khan, M. U., Roy, N. C., Islam, R., Huq, I. and Stoll, B. (1985) 'Fourteen years of shigellosis in Dhaka: an epidemiological analysis', *Int J Epidemiol*, 14(4), pp. 607–613.
- Khan, W. a, Griffiths, J. K. and Bennish, M. L. (2013) 'Gastrointestinal and extra-intestinal manifestations of childhood shigellosis in a region where all four species of Shigella are endemic.', *PloS one*, 8(5), p. e64097. doi: 10.1371/journal.pone.0064097.
- Kim, J. S., Kim, J. J., Kim, S. J., Jeon, S., Seo, K. Y., Choi, J., Kim, N., Hong, S., Chung, G. T., Yoo, C., Kim, Y., Cheun, H. Il, Bae, G., Yeo, Y., Ha, G. and Choi, M. (2015) 'Shigella sonnei Associated with Travel to Vietnam, Republic of Korea', *Emerging Infectious Diseases*, 21(7), pp. 1247–1250.
- Kim, M., Ogawa, M., Fujita, Y., Yoshikawa, Y., Nagai, T., Koyama, T., Nagai, S., Lange, A., Fässler, R. and Sasakawa, C. (2009) 'Bacteria hijack integrin-linked kinase to stabilize focal adhesions and block cell detachment.', *Nature*, 459(7246), pp. 578–82. doi: 10.1038/nature07952.
- Kim, S. R. and Komano, T. (1997) 'The plasmid R64 thin pilus identified as a type IV pilus', *Journal of Bacteriology*, 179(11), pp. 3594–3603.
- Knirel, Y. A., Sun, Q., Senchenkova, S. N., Perepelov, A. V, Shashkov, A. S. and Xu, J. (2015) 'O Antigen Modifications Providing Antigenic Diversity of Shigella flexneri and Underlying Genetic Mechanisms', *Biochemistry (Moscow)*, 80(7).
- Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., Angenent, L. T. and Ley, R. E. (2011) 'Succession of microbial consortia in the developing infant gut microbiome.', *Proceedings of the National Academy of*

Sciences, 108(suppl 1), pp. 4578–4585. doi: 10.1073/pnas.1000081107.

- Koren, O., Knights, D., Gonzalez, A., Waldron, L., Segata, N., Knight, R., Huttenhower, C. and Ley, R. E. (2013) 'A Guide to Enterotypes across the Human Body: Meta-Analysis of Microbial Community Structures in Human Microbiome Datasets', *PLoS Computational Biology*, 9(1), p. e1002863. doi: 10.1371/journal.pcbi.1002863.
- Kostic, A. D., Gevers, D., Pedamallu, C. S., Kostic, A. D., Gevers, D., Pedamallu, C. S., Michaud, M., Duke, F., Earl, A. M., Ojesina, A. I., Jung, J., Bass, A. J., Liu, C., Shivdasani, R. A., Ogino, S., Tabernero, J., Birren, B. W., Huttenhower, C., Garrett, W. S. and Meyerson, M. (2012) 'Genomic analysis identifies association of Fusobacterium with colorectal carcinoma', *Genome Res*, 22, pp. 292–298. doi: 10.1101/gr.126573.111.
- Kotloff, K. L., Blackwelder, W. C., Nasrin, D., Nataro, J. P., Farag, T. H., van Eijk, A., Adegbola, R. a, Alonso, P. L., Breiman, R. F., Faruque, A. S. G., Saha, D., Sow, S. O., Sur, D., Zaidi, A. K. M., Biswas, K., Panchalingam, S., Clemens, J. D., Cohen, D., Glass, R. I., Mintz, E. D., Sommerfelt, H. and Levine, M. M. (2012) 'The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiologic and clinical methods of the case/control study.', *Clinical infectious diseases*, 55 Suppl 4(Suppl 4), pp. S232-45. doi: 10.1093/cid/cis753.
- Kotloff, K. L., Nataro, J. P., Blackwelder, W. C., Nasrin, D., Farag, T. H., Panchalingam, S., Wu, Y., Sow, S. O., Sur, D., Breiman, R. F., Faruque, A. S., Zaidi, A. K., Saha, D., Alonso, P. L., Tamboura, B., Sanogo, D., Onwuchekwa, U., Manna, B., Ramamurthy, T., Kanungo, S., Ochieng, J. B., Omore, R., Oundo, J. O., Hossain, A., Das, S. K., Ahmed, S., Qureshi, S., Quadri, F., Adegbola, R. a, Antonio, M., Hossain, M. J., Akinsola, A., Mandomando, I., Nhampossa, T., Acácio, S., Biswas, K., O'Reilly, C. E., Mintz, E. D., Berkeley, L. Y., Muhsen, K., Sommerfelt, H., Robins-Browne, R. M. and Levine, M. M. (2013) 'Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study.', *Lancet*, 382(9888), pp. 209–22. doi: 10.1016/S0140-6736(13)60844-2.

- Kotloff, K. L., Winickoff, J. P., Ivanoff, B., Clemens, J. D., Swerdlow, D. L., Sansonetti, P. J., Adak, G. K. and Levine, M. M. (1999) 'Global burden of Shigella infections: implications for vaccine development and implementation of control strategies', *Bulletin of the World Health Organization*. 1999/10/12, 77(8), pp. 651–666. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10516787.
- Kotwani, A., Chaudhury, R. R. and Holloway, K. (2012) 'Antibiotic-prescribing practice of primary care prescribers for acute diarrhea in New Delhi, India', *Value in Health*, 15, pp. S16–S119.
- Kouse, A. B., Righetti, F., Kortmann, J., Narberhaus, F. and Murphy, E. R. (2013) 'RNA-mediated thermoregulation of iron-acquisition genes in Shigella dysenteriae and pathogenic Escherichia coli.', *PLoS ONE*, 8(5), p. e63781. doi: 10.1371/journal.pone.0063781.
- Kozyreva, V. K., Jospin, G., Greninger, A. L., Watt, J. P., Eisen, J. A. and Chaturvedi, V. (2016) 'Recent Outbreaks of Shigellosis in California Caused by Two Distinct Populations of Shigella sonnei with either Increased Virulence or Fluoroquinolone Resistance', *mSphere*, 1(6), pp. 1–18. doi: 10.1128/mSphere.00344-16.Editor.
- Krieg, N. R., Staley, J. T., Brown, D. R., Hedlund, B. P., Paster, B. J., Ward, N. L., Ludwig, W. and Whitman, W. B. (2010) 'Volume Four The Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes', in Brown, D. R., Hedlund, B. P., Paster, B. J., Ward, N. L., Ludwig, W., and Whitman, W. B. (eds) *Bergey's Manual of Systematic Bacteriology*. Second Edi. New York,: Springer, p. 949. doi: 10.1007/978-0-387-68572-4.
- Kryazhimskiy, S. and Plotkin, J. B. (2008) 'The population genetics of dN/dS', *PLoS Genetics*, 4(12). doi: 10.1371/journal.pgen.1000304.
- Künne, C., Billion, A., Mshana, S. E., Schmiedel, J., Domann, E., Hossain, H., Hain, T., Imirzalioglu, C. and Chakraborty, T. (2012) 'Complete sequences of plasmids from the hemolytic-uremic syndrome-associated Escherichia coli strain HUSEC41', *Journal of Bacteriology*, 194, pp. 532–533. doi: 10.1128/JB.06368-11.
- Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J. and

Bonneau, R. A. (2015) ‘Sparse and Compositionally Robust Inference of Microbial Ecological Networks’, *PLOS Computational Biology*, 11(5), p. e1004226. doi: 10.1371/journal.pcbi.1004226.

- Lacher, D. W., Steinsland, H., Blank, T. E., Donnenberg, M. S. and Whittam, T. S. (2007) ‘Molecular evolution of typical enteropathogenic Escherichia coli: clonal analysis by multilocus sequence typing and virulence gene allelic profiling’, *Journal of bacteriology*. 2006/11/14, 189(2), pp. 342–350. doi: 10.1128/JB.01472-06.
- Lafont, F., Nhieu, G. T. Van, Hanada, K., Sansonetti, P. and Goot, F. G. Van Der (2002) ‘Initial steps of Shigella infection depend on the cholesterol / sphingolipid raft-mediated CD44 ± IpaB interaction’, *The EMBO Journal*, 21(17), pp. 4449–4457.
- Lan, R., Alles, M. C., Donohoe, K., Marina, B., Reeves, P. R. and Martinez, M. B. (2004) ‘Molecular Evolutionary Relationships of Enteroinvasive Escherichia coli and Shigella spp .’, *Infection and Immunity*. doi: 10.1128/IAI.72.9.5080.
- Lan, R., Lumb, B., Ryan, D. and Reeves, P. R. (2001) ‘Molecular Evolution of Large Virulence Plasmid in Shigella Clones and Enteroinvasive Escherichia coli’, *Infection and Immunity*, 69(10), pp. 6303–6309. doi: 10.1128/IAI.69.10.6303.
- Lan, R. and Reeves, P. R. (2002) ‘Escherichia coli in disguise: molecular origins of Shigella’, *Microbes and infection / Institut Pasteur*. 2002/10/04, 4(11), pp. 1125–1132. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12361912.
- Landini, P., Egli, T., Wolf, J. and Lacour, S. (2014) ‘sigmaS , a major player in the response to environmental stresses in Escherichia coli : role , regulation and mechanisms of promoter recognition’, *Environmental Microbiology Reports*, 6, pp. 1–13. doi: 10.1111/1758-2229.12112.
- Lapierre, M., Blin, C., Lambert, A., Achaz, G. and Rocha, E. P. C. (2016) ‘The Impact of Selection , Gene Conversion , and Biased Sampling on the Assessment of Microbial Demography’, *Molecular Biology and Evolution*, 33(7), pp. 1711–1725. doi: 10.1093/molbev/msw048.
- De Lappe, N., Connor, J. O., Garvey, P., Mckeown, P. and Cormican, M. (2015) ‘Ciprofloxacin-Resistant Shigella sonnei Associated with Travel to India’, *Emerging Infectious Diseases*, 21(5), pp. 2013–2015.
- Laxminarayan, R. and Chaudhury, R. R. (2016) ‘Antibiotic Resistance in India:

Drivers and Opportunities for Action', *PLoS Medicine*, 13(3), pp. 1–7. doi: 10.1371/journal.pmed.1001974.

- Levesque, C., Lamothe, J. and Frenette, M. (2003) 'Coaggregation of Streptococcus salivarius with periodontopathogens: evidence for involvement of fimbriae in the interaction with Prevotella intermedia', *Oral Microbiol Immunol*, 18(5), pp. 333–337. doi: 085 [pii].
- Li, H. (2013) 'Aligning sequence reads , clone sequences and assembly contigs with BWA-MEM', *bioarxiv*, 0(0), pp. 1–3.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) 'The Sequence Alignment/Map format and SAMtools.', *Bioinformatics (Oxford, England)*, 25(16), pp. 2078–9. doi: 10.1093/bioinformatics/btp352.
- Liang, B., Roberts, A. P., Xu, X., Yang, C., Yang, X., Wang, J., Yi, S., Li, Y., Ma, Q., Wu, F., Qiu, S. and Song, H. (2018) 'Transferable Plasmid-Borne mcr-1 in a Colistin Resistant Shigella flexneri Isolate', *mBio*, 84(8), pp. 1–10.
- Lieberman, T. D., Michel, J.-B., Aingaran, M., Potter-Bynoe, G., Roux, D., Davis, M. R., Skurnik, D., Leiby, N., LiPuma, J. J., Goldberg, J. B., McAdam, A. J., Priebe, G. P. and Kishony, R. (2011) 'Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes', *Nature Genetics*. Nature Publishing Group, 43(12), pp. 1275–1280. doi: 10.1038/ng.997.
- Lindsay, B., Ochieng, J. B., Ikumapayi, U. N., Toure, A., Ahmed, D., Li, S., Panchalingam, S., Levine, M. M., Kotloff, K., Rasko, D. A., Morris, C. R., Juma, J., Fields, B. S., Dione, M., Malle, D., Becker, S. M., Houpt, E. R., Nataro, J. P., Sommerfelt, H., Pop, M., Oundo, J., Antonio, M., Hossain, A., Tamboura, B. and Stine, O. C. (2013) 'Quantitative PCR for detection of Shigella improves ascertainment of Shigella burden in children with moderate-to-severe diarrhea in low-income countries', *Journal of Clinical Microbiology*, 51(6), pp. 1740–1746. doi: 10.1128/JCM.02713-12.
- Liu, Z., Vargas-rodriguez, O., Goto, Y., Maria, E., Ribas, L., Pouplana, D. and Suga, H. (2015) 'Homologous trans -editing factors with broad tRNA specificity prevent mistranslation caused by serine / threonine misactivation', *PNAS*, 112(19), pp. 6027–

6032. doi: 10.1073/pnas.1423664112.

- Livio, S., Strockbine, N. a, Panchalingam, S., Tennant, S. M., Barry, E. M., Marohn, M. E., Antonio, M., Hossain, A., Mandomando, I., Ochieng, J. B., Oundo, J. O., Qureshi, S., Ramamurthy, T., Tamboura, B., Adegbola, R. a, Hossain, M. J., Saha, D., Sen, S., Faruque, A. S. G., Alonso, P. L., Breiman, R. F., Zaidi, A. K. M., Sur, D., Sow, S. O., Berkeley, L. Y., O'Reilly, C. E., Mintz, E. D., Biswas, K., Cohen, D., Farag, T. H., Nasrin, D., Wu, Y., Blackwelder, W. C., Kotloff, K. L., Nataro, J. P. and Levine, M. M. (2014) 'Shigella Isolates From the Global Enteric Multicenter Study Inform Vaccine Development.', *Clinical infectious diseases*, 59(7), pp. 933–941. doi: 10.1093/cid/ciu468.
- Lomovskaya, O., Lewis, K. and Matin, A. (1995) 'EmrR Is a Negative Regulator of the Escherichia coli Multidrug Resistance pump EmrAB', *Journal of Bacteriology*, 177(9), pp. 2328–2334.
- Lopez, C. A., Kingsbury, D. D., Velazquez, E. M. and Baumler, A. J. (2014) 'Collateral damage: Microbiota-derived metabolites and immune function in the antibiotic era', *Cell Host and Microbe*, 16(2), pp. 156–163. doi: 10.1016/j.chom.2014.07.009.
- Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), p. 550. doi: 10.1186/s13059-014-0550-8.
- Lozupone, C. A., Hamady, M., Kelley, S. T. and Knight, R. (2007) 'Quantitative and qualitative beta-diversity measures lead to different insights into factors that structure microbial communities', *Applied and Environmental Microbiology*, 73(5), pp. 1576–1585. doi: 10.1128/AEM.01996-06.
- Luck, S. N., Turner, S. A., Rajakumar, K., Sakellaris, H. and Adler, B. (2001) 'Ferric dicitrate transport system (Fec) of Shigella flexneri 2a YSH6000 is encoded on a novel pathogenicity island carrying multiple antibiotic resistance genes', *Infection and immunity*. 2001/09/13, 69(10), pp. 6012–6021. doi: 10.1128/IAI.69.10.6012-6021.2001.
- Luo, X., Sun, Q., Lan, R., Wang, J., Li, Z., Xia, S., Zhang, J., Wang, Y., Jin, D., Yuan, X., Yu, B., Cui, Z. and Xu, J. (2012) 'Emergence of a novel Shigella flexneri

serotype 1d in China', *Diagnostic microbiology and infectious disease*. 2012/08/04, 74(3), pp. 316–319. doi: 10.1016/j.diagmicrobio.2012.06.022.

- MacIntyre, D. A., Chandiramani, M., Lee, Y. S., Kindinger, L., Smith, A., Angelopoulos, N., Lehne, B., Arulkumaran, S., Brown, R., Teoh, T. G., Holmes, E., Nicoholson, J. K., Marchesi, J. R. and Bennett, P. R. (2015) 'The vaginal microbiome during pregnancy and the postpartum period in a European population.', *Scientific reports*, 5(Cst Iv), p. 8988. doi: 10.1038/srep08988.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. and Hornik, K. (2016) 'cluster: Cluster Analysis Basics and Extensions. R package version 2.0.4'.
- Makino, S., Sasakawa, C. and Yoshikawa, M. (1988) 'Genetic relatedness of the basic replicon of the virulence plasmid in shigellae and enteroinvasive Escherichia coli.', *Microb Pathog*, 5(4), pp. 267–274.
- Marcusson, L. L., Frimodt-Møller, N. and Hughes, D. (2009) 'Interplay in the selection of fluoroquinolone resistance and bacterial fitness', *PLoS Pathogens*, 5(8). doi: 10.1371/journal.ppat.1000541.
- Marfin, a a, Moore, J., Collins, C., Biellik, R., Kattel, U., Toole, M. J. and Moore, P. S. (1994) 'Infectious disease surveillance during emergency relief to Bhutanese refugees in Nepal.', *JAMA : the journal of the American Medical Association*, 272(5), pp. 377–381. doi: 10.1001/jama.272.5.377.
- Marincs, F., Manfield, I. W., Stead, J. A., Mcdowall, K. J. and Stockley, P. G. (2006) 'Transcript analysis reveals an extended regulon and the importance of protein–protein co-operativity for the Escherichia coli methionine repressor', *Biochem. J*, 396, pp. 227–234. doi: 10.1042/BJ20060021.
- Martinez-Becerra, F. J., Chen, X., Dickenson, N. E., Choudhari, S. P., Harrison, K., Clements, J. D., Picking, W. D., Van De Verg, L. L., Walker, R. I. and Picking, W. L. (2013) 'Characterization of a novel fusion protein from IpaB and IpaD of Shigella spp. and its potential as a pan-Shigella vaccine.', *Infection and immunity*, 81(12), pp. 4470–7. doi: 10.1128/IAI.00859-13.
- Martínez-García, E., Nikel, P. I., Chavarría, M. and de Lorenzo, V. (2014) 'The metabolic cost of flagellar motion in Pseudomonas putidaKT2440', *Environmental Microbiology*, 16(1), pp. 291–303. doi: 10.1111/1462-2920.12309.

- Marx, V. and Fleischer, K. (2010) *Internal Migration: Opportunities and challenges for socio-economic development in Vietnam*. doi: 10.1038/1811574d0.
- Maurelli, A. T., Fernandez, R. E., Bloch, C. A., Rode, C. K. and Fasano, A. (1998) '"Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of Shigella spp. and enteroinvasive Escherichia coli', *Proceedings of the National Academy of Sciences of the United States of America*. 1998/05/09, 95(7), pp. 3943–3948. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9520472.
- Mavris, M., Manning, P. a and Morona, R. (1997) 'Mechanism of bacteriophage SfII-mediated serotype conversion in Shigella flexneri.', *Molecular microbiology*, 26(5), pp. 939–50. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9426131.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. and DePristo, M. (2010) 'The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data', *Genome Research*, 20, pp. 1297–303. doi: 10.1101/gr.107524.110.20.
- McMurdie, P. J. and Holmes, S. (2013) 'phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data.', *PloS one*, 8(4), p. e61217. doi: 10.1371/journal.pone.0061217.
- McMurdie, P. J. and Holmes, S. (2014) 'Waste not, want not: why rarefying microbiome data is inadmissible.', *PLoS computational biology*, 10(4), p. e1003531. doi: 10.1371/journal.pcbi.1003531.
- McVicker, G. and Tang, C. M. (2016) 'Deletion of toxin–antitoxin systems in the evolution of Shigella sonnei as a host-adapted pathogen', *Nature Microbiology*. Nature Publishing Group, 2(November), p. 16204. doi: 10.1038/nmicrobiol.2016.204.
- Meng, C. Y., Smith, B. L., Bodhidatta, L., Richard, S. A., Vansith, K., Thy, B., Srijan, A., Serichantalergs, O. and Mason, C. J. (2011) 'Etiology of diarrhea in young children and patterns of antibiotic resistance in Cambodia', *Pediatric Infectious Disease Journal*, 30(4), pp. 331–335. doi: 10.1097/INF.0b013e3181fb6f82.
- Miller, C. S., Baker, B. J., Thomas, B. C., Singer, S. W. and Banfield, J. F. (2011)

'EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data', *Genome Biology*. BioMed Central Ltd, 12(5), p. R44. doi: 10.1186/gb-2011-12-5-r44.

- Minin, V. N., Bloomquist, E. W. and Suchard, M. A. (2008) 'Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics', *Molecular Biology and Evolution*, 25(7), pp. 1459–1471. doi: 10.1093/molbev/msn090.
- Mohan, K., Hibbert, M., Rooney, G., Canvin, M., Childs, T., Jenkins, C., Simms, I., Kirwan, P., Delpech, V., Yin, Z., Hughes, G. and Field, N. (2017) 'What is the overlap between HIV and shigellosis epidemics in England: further evidence of MSM transmission?', *Sexually Transmitted Infections*, p. sextrans-2016-052962. doi: 10.1136/sextrans-2016-052962.
- Monira, S., Hoq, M. M., Chowdhury, A. K. A., Suau, A., Magne, F., Endtz, Hp., Alam, M., Rahman, M., Pochart, P., Desjeux, J.-F. and Alam, N. H. (2010) 'Short-chain fatty acids and commensal microbiota in the faeces of severely malnourished children with cholera rehydrated with three different carbohydrates', *European Journal of Clinical Nutrition*. Nature Publishing Group, 64(10), pp. 1116–1124. doi: 10.1038/ejcn.2010.123.
- Monira, S., Nakamura, S., Gotoh, K., Izutsu, K., Watanabe, H., Alam, N. H., Nakaya, T., Horii, T., Ali, S. I., Iida, T. and Alam, M. (2013) 'Metagenomic profile of gut microbiota in children during cholera and recovery.', *Gut pathogens*, 5(1), p. 1. doi: 10.1186/1757-4749-5-1.
- Monk, J. M., Charusanti, P., Aziz, R. K., Lerman, J. A., Premyodhin, N. and Orth, J. D. (2013) 'Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1307797110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1307797110.
- Moratalla, A., Caparrós, E., Juanola, O., Portune, K., Puig-Kröger, A., Estrada-Capetillo, L., Bellot, P., Gómez-Hurtado, I., Piñero, P., Zapater, P., González-Navajas, J. M., Such, J., Sanz, Y. and Francés, R. (2015) 'Bifidobacterium pseudocatenulatum CECT7765 induces an M2 anti-inflammatory transition in

macrophages from patients with cirrhosis', *Journal of Hepatology*. European Association for the Study of the Liver, 64, pp. 135–145. doi: 10.1016/j.jhep.2015.08.020.

- Moratalla, A., Gómez-Hurtado, I., Moya-Pérez, Á., Zapater, P., Peiró, G., González-Navajas, J. M., Gómez Del Pulgar, E. M., Such, J., Sanz, Y. and Francés, R. (2016) 'Bifidobacterium pseudocatenulatum CECT7765 promotes a TLR2-dependent anti-inflammatory response in intestinal lymphocytes from mice with cirrhosis', *European Journal of Nutrition*, 55(1), pp. 197–206. doi: 10.1007/s00394-015-0837-x.
- Moratalla, A., Gómez-Hurtado, I., Santacruz, A., Moya, Á., Peiró, G., Zapater, P., González-Navajas, J. M., Giménez, P., Such, J., Sanz, Y. and Francés, R. (2014) 'Protective effect of Bifidobacterium pseudocatenulatum CECT7765 against induced bacterial antigen translocation in experimental cirrhosis', *Liver International*, 34(11), pp. 850–858. doi: 10.1111/liv.12380.
- Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V, Reyes, J. a, Shah, S. a, LeLeiko, N., Snapper, S. B., Bousvaros, A., Korzenik, J., Sands, B. E., Xavier, R. J. and Huttenhower, C. (2012) 'Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment.', *Genome biology*, 13(9), p. R79. doi: 10.1186/gb-2012-13-9-r79.
- Mounier, J., Boncompain, G., Senerovic, L., Lagache, T., Chrétien, F., Perez, F., Kolbe, M., Olivo-Marin, J.-C., Sansonetti, P. J. and Sauvonnet, N. (2012) 'Shigella effector IpaB-induced cholesterol relocation disrupts the Golgi complex and recycling network to inhibit host cell secretion.', *Cell host & microbe*, 12(3), pp. 381–9. doi: 10.1016/j.chom.2012.07.010.
- Mounier, J., Popoff, M. R., Enninga, J., Frame, M. C., Sansonetti, P. J. and Van Nhieu, G. T. (2009) 'The IpaC carboxyterminal effector domain mediates Src-dependent actin polymerization during Shigella invasion of epithelial cells.', *PLoS pathogens*, 5(1), p. e1000271. doi: 10.1371/journal.ppat.1000271.
- Moya-Pérez, A., Neef, A. and Sanz, Y. (2015) 'Bifidobacterium pseudocatenulatum CECT 7765 Reduces Obesity-Associated Inflammation by Restoring the Lymphocyte-Macrophage Balance and Gut Microbiota Structure in High-Fat Diet-

Fed Mice', *Plos One*, 10(7), p. e0126976. doi: 10.1371/journal.pone.0126976.

- My, P. V. T., Rabaa, M. a., Vinh, H., Holmes, E. C., Hoang, N. V. M., Vinh, N. T., Phuong, L. T., Tham, N. T., Bay, P. V. B., Campbell, J. I., Farrar, J. and Baker, S. (2011) 'The emergence of rotavirus G12 and the prevalence of enteric viruses in hospitalized pediatric diarrheal patients in Southern Vietnam', *American Journal of Tropical Medicine and Hygiene*, 85(4), pp. 768–775. doi: 10.4269/ajtmh.2011.11-0364.
- My, P. V. T., Thompson, C. N., Phuc, H. Le, Tuyet, P. T. N., Vinh, H., Nguyen, V. M. H., Minh, P. Van, Vinh, N. T., Thuy, C. T., Tran, T. T. N., Thi, N. H. T., Campbell, J. I., Chinh, N. T., Thuong, T. C., Tuan, H. M., Farrar, J. and Baker, S. (2013) 'Endemic norovirus infections in children, Ho Chi Minh City, Vietnam, 2009-2010.', *Emerging infectious diseases*, 19(6), pp. 29–32. doi: 10.3201/eid1906.111862.
- Naheed, A., Kalluri, P., Talukder, K. A., Faruque, A. S. G., Khatun, F., Nair, G. B., Mintz, E. D. and Breiman, R. F. (2004) 'Fluoroquinolone-resistant Shigella dysenteriae type 1 in northeastern Bangladesh', *Lancet Infectious Diseases*, pp. 607–608. doi: 10.1016/S1473-3099(04)01143-0.
- Nakata, N., Tobe, T. and Fukuda, I. (1993) 'The absence of a surface protease , OmpT , determines the intercellular spreading ability of Shigella:the relationship between the ompT and kcpA loci', *Molecular Microbiology*, 9(3), pp. 459–468. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2958.1993.tb01707.x/abstract (Accessed: 9 January 2015).
- Nandy, S., Dutta, S., Ghosh, S., Ganai, A., Rajahamsan, J., Theodore, R. B. J. and Sheikh, N. K. (2011) 'Foodborneassociated Shigella sonnei, India, 2009 and 2010', *Emerging Infectious Diseases*, 17(11), pp. 2072–2073. doi: 10.3201/eid1502.080988.
- Naseribafrouei, A., Hestad, K., Avershina, E., Sekelja, M., Linløkken, A., Wilson, R. and Rudi, K. (2014) 'Correlation between the human fecal microbiota and depression', *Neurogastroenterology and Motility*, 26(8), pp. 1155–1162. doi: 10.1111/nmo.12378.
- Nataro, J. P., Seriwatana, J., Fasano, A., Maneval, D. R., Guers, L. D., Noriega, F., Dubovsky, F., Levine, M. M. and Morris, J. G. (1995) 'Identification and Cloning of

a Novel Plasmid-Encoded Enterotoxin of Enteroinvasive Escherichia coli and Shigella Strains', *Infection and immunity*, 63(12), pp. 4721–4728.

- Naves, P., Del Prado, G., Huelves, L., Gracia, M., Ruiz, V., Blanco, J., Rodr??guez-Cerrato, V., Ponte, M. C. and Soriano, F. (2008) 'Measurement of biofilm formation by clinical isolates of Escherichia coli is method-dependent', *Journal of Applied Microbiology*, 105(2), pp. 585–590. doi: 10.1111/j.1365-2672.2008.03791.x.
- Nayfach, S., Rodriguez-Mueller, B., Garud, N. and Pollard, K. S. (2016) 'An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography', *Genome Research*, 26(11), pp. 1612–1625. doi: 10.1101/gr.201863.115.
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. and Minh, B. Q. (2015) 'IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies', *Molecular Biology and Evolution*, 32(1), pp. 268–274. doi: 10.1093/molbev/msu300.
- Niebuhr, K., Giuriato, S. and Pedron, T. (2002) 'Conversion of PtdIns (4, 5) P2 into PtdIns (5) P by the S. flexneri effector IpgD reorganizes host cell morphology', *The EMBO Journal*, 21(19), pp. 5069–5078. Available at: http://onlinelibrary.wiley.com/doi/10.1093/emboj/cdf522/full (Accessed: 9 January 2015).
- Nielsen, R. and Yang, Z. (1998) 'Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene', *Genetics*, 148(3), pp. 929–936.
- Niyogi, S. K. (2005) 'Shigellosis', *Journal of microbiology*. 2005/05/10, 43(2), pp. 133–143. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15880088.
- Niyogi, S. K. (2007) 'Increasing antimicrobial resistance--an emerging problem in the treatment of shigellosis', *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*. 2007/10/24, 13(12), pp. 1141–1143. doi: 10.1111/j.1469-0691.2007.01829.x.
- Njamkepo, E., Fawal, N., Tran-Dien, A., Hawkey, J., Strockbine, N., Jenkins, C., Talukder, K. A., Bercion, R., Kuleshov, K., Kolínská, R., Russell, J. E., Kaftyreva,

L., Accou-Demartin, M., Karas, A., Vandenberg, O., Mather, A. E., Mason, C. J., Page, A. J., Ramamurthy, T., Bizet, C., Gamian, A., Carle, I., Sow, A. G., Bouchier, C., Wester, A. L., Lejay-Collin, M., Fonkoua, M.-C., Hello, S. Le, Blaser, M. J., Jernberg, C., Ruckly, C., Mérens, A., Page, A.-L., Aslett, M., Roggentin, P., Fruth, A., Denamur, E., Venkatesan, M., Bercovier, H., Bodhidatta, L., Chiou, C.-S., Clermont, D., Colonna, B., Egorova, S., Pazhani, G. P., Ezernitchi, A. V., Guigon, G., Harris, S. R., Izumiya, H., Korzeniowska-Kowal, A., Lutyńska, A., Gouali, M., Grimont, F., Langendorf, C., Marejková, M., Peterson, L. A. M., Perez-Perez, G., Ngandjio, A., Podkolzin, A., Souche, E., Makarova, M., Shipulin, G. A., Ye, C., Žemličková, H., Herpay, M., Grimont, P. A. D., Parkhill, J., Sansonetti, P., Holt, K. E., Brisse, S., Thomson, N. R. and Weill, F.-X. (2016) 'Global phylogeography and evolutionary history of Shigella dysenteriae type 1', *Nature Microbiology*, (March), p. 16027. doi: 10.1038/nmicrobiol.2016.27.

- Noriega, F. R., Liao, F. M., Maneval, D. R., Ren, S., Formal, S. B. and Levine, M. M. (1999) 'Strategy for cross-protection among Shigella flexneri serotypes', *Infection and immunity*. 1999/01/23, 67(2), pp. 782–788. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9916090.
- Nygren, B. L., Schilling, K. A., Blanton, E. M., Silk, B. J., Cole, D. J. and Mintz, E. D. (2013) 'Foodborne outbreaks of shigellosis in the USA, 1998-2008', *Epidemiology and infection*. 2012/03/01, 141(2), pp. 233–241. doi: 10.1017/S0950268812000222.
- Ochman, H. and Davalos, L. M. (2006) 'The nature and dynamics of bacterial genomes.', *Science (New York, N.Y.)*, 311(5768), pp. 1730–3. doi: 10.1126/science.1119966.
- Oksanen, J., Blanchet, G., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Simpson, G., Solymos, P., Stevens, M. and Wagner, H. (2016) 'vegan: Community Ecology Package. R package version 2.3-5'.
- Ong, S. H., Kukkillaya, V. U., Wilm, A., Lay, C., Ho, E. X. P., Low, L., Hibberd, M. L. and Nagarajan, N. (2013) 'Species Identification and Profiling of Complex Microbial Communities Using Shotgun Illumina Sequencing of 16S rRNA Amplicon Sequences', *PLoS ONE*, 8(4), pp. 1–8. doi: 10.1371/journal.pone.0060811.

- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A. and Parkhill, J. (2015) 'Roary: Rapid large-scale prokaryote pan genome analysis', *Bioinformatics*, 31(22), pp. 3691–3693. doi: 10.1093/bioinformatics/btv421.
- Page, A. J., De Silva, N., Hunt, M., Quail, M. A., Parkhill, J., Otto, T. D., Harris, S. R. and Keane, J. A. (2016) 'Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data', *Microbial Genomics*, 2(8), pp. 1–7. doi: 10.1099/mgen.0.000083.
- Pal, S. C. (1984) 'Epidemic bacillary dysentery in West Bengal, India, 1984.', *The Lancet*, 323(8392), p. 1462. Available at: https://auth.lib.unc.edu/ezproxy_auth.php?url=http://search.ebscohost.com/login.aspx?direct=true&db=lhh&AN=19842016757&site=ehost-live&scope=site.
- Park, H., Valencia-Gallardo, C., Sharff, A., Van Nhieu, G. T. and Izard, T. (2011) 'Novel vinculin binding site of the IpaA invasin of Shigella', *Journal of Biological Chemistry*, 286, pp. 23214–23221. doi: 10.1074/jbc.M110.184283.
- Parkhill, J., Wren, B. W., Thomson, N. R., Titball, R. W., Holden, M. T., Prentice, M. B., Sebaihia, M., James, K. D., Churcher, C., Mungall, K. L., Baker, S., Basham, D., Bentley, S. D., Brooks, K., Cerdeño-Tárraga, a M., Chillingworth, T., Cronin, a, Davies, R. M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Karlyshev, a V, Leather, S., Moule, S., Oyston, P. C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrell, B. G. (2001) 'Genome sequence of Yersinia pestis, the causative agent of plague.', *Nature*, 413(6855), pp. 523–527. doi: 10.1038/35097083.
- Pazhani, G. P., Niyogi, S. K., Singh, A. K., Sen, B., Taneja, N., Kundu, M., Yamasaki, S. and Ramamurthy, T. (2008) 'Molecular characterization of multidrug-resistant Shigella species isolated from epidemic and endemic cases of shigellosis in India', *Journal of Medical Microbiology*, 57(7), pp. 856–863. doi: 10.1099/jmm.0.2008/000521-0.
- Pedersen, H. K., Gudmundsdottir, V., Nielsen, H. B., Hyotylainen, T., Nielsen, T., Jensen, B. A. H., Forslund, K., Hildebrand, F., Prifti, E., Falony, G., Le Chatelier, E., Levenez, F., Doré, J., Mattila, I., Plichta, D. R., Pöhö, P., Hellgren, L. I., Arumugam,

M., Sunagawa, S., Vieira-Silva, S., Jørgensen, T., Holm, J. B., Trošt, K., Consortium, M., Kristiansen, K., Brix, S., Raes, J., Wang, J., Hansen, T., Bork, P., Brunak, S., Oresic, M., Ehrlich, S. D. and Pedersen, O. (2016) 'Human gut microbes impact host serum metabolome and insulin sensitivity.', *Nature*. Nature Publishing Group, 535(7612), pp. 376–381. doi: 10.1038/nature18646.

- Pendaries, C., Tronchère, H., Arbibe, L., Mounier, J., Gozani, O., Cantley, L., Fry, M. J., Gaits-Iacovoni, F., Sansonetti, P. J. and Payrastre, B. (2006) 'PtdIns5P activates the host cell PI3-kinase/Akt pathway during Shigella flexneri infection.', *The EMBO journal*, 25(5), pp. 1024–34. doi: 10.1038/sj.emboj.7601001.
- Phan, D. and Coxhead, I. (2010) 'Inter-provincial migration and inequality during Vietnam's transition', *Journal of Development Economics*. Elsevier B.V., 91(1), pp. 100–112. doi: 10.1016/j.jdeveco.2009.06.008.
- Ponsting, H. and Ning, Z. (2010) 'SMALT - A New Mapper for DNA Sequencing Reads', *F1000Posters*, p. 1. doi: 10.7490/F1000RESEARCH.327.1.
- Pop, M., Walker, A. W., Paulson, J., Lindsay, B., Antonio, M., Hossain, M., Oundo, J., Tamboura, B., Mai, V., Astrovskaya, I., Bravo, H., Rance, R., Stares, M., Levine, M. M., Panchalingam, S., Kotloff, K., Ikumapayi, U. N., Ebruke, C., Adeyemi, M., Ahmed, D., Ahmed, F., Alam, M., Amin, R., Siddiqui, S., Ochieng, J. B., Ouma, E., Juma, J., Mailu, E., Omore, R., Morris, J., Breiman, R. F., Saha, D., Parkhill, J., Nataro, J. P. and Stine, O. (2014) 'Diarrhea in young children from low-income countries leads to large-scale alterations in intestinal microbiota composition', *Genome Biology*, 15(6), p. R76. doi: 10.1186/gb-2014-15-6-r76.
- Portrait, V., Cottenceau, G. and Pons, a M. (2000) 'A Fusobacterium mortiferum strain produces a bacteriocin-like substance(s) inhibiting Salmonella enteritidis.', *Letters in applied microbiology*, 31(2), pp. 115–7. doi: 10.1046/j.1365-2672.2000.00774.x.
- Posada, D. (2008) 'jModelTest: phylogenetic model averaging.', *Molecular biology and evolution*, 25(7), pp. 1253–6. doi: 10.1093/molbev/msn083.
- Price, M. N., Dehal, P. S. and Arkin, A. P. (2010) 'FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments', *PLoS ONE*, 5(3), p. e9490. doi: 10.1371/journal.pone.0009490.

- Prigent-combaret, C., Vidal, O., Dorel, C. and Lejeune, P. (1999) 'Abiotic Surface Sensing and Biofilm-Dependent Regulation of Gene Expression in Escherichia coli Abiotic Surface Sensing and Biofilm-Dependent Regulation of Gene Expression in Escherichia coli', 181(19), pp. 5993–6002.
- Prosseda, G., Di Martino, M. L., Campilongo, R., Fioravanti, R., Micheli, G., Casalino, M. and Colonna, B. (2012) 'Shedding of genes that interfere with the pathogenic lifestyle: the Shigella model', *Research in microbiology*. 2012/07/25, 163(6–7), pp. 399–406. doi: 10.1016/j.resmic.2012.07.004.
- Prunier, A. L., Schuch, R., Fernandez, R. E., Mumy, K. L., Kohler, H., McCormick, B. A. and Maurelli, A. T. (2007) 'nadA and nadB of Shigella flexneri 5a are antivirulence loci responsible for the synthesis of quinolinate, a small molecule inhibitor of Shigella pathogenicity', *Microbiology*. 2007/06/30, 153(Pt 7), pp. 2363–2372. doi: 10.1099/mic.0.2007/006916-0.
- Pu, X., Pan, J., Wang, H., Zhang, W., Huang, Z. and Gu, Y. (2009) 'Characterization of fluoroquinolone-resistant Shigella flexneri in Hangzhou area of China', *Journal of Antimicrobial Chemotherapy*, (October 2017), pp. 917–920. doi: 10.1093/jac/dkp087.
- Pu, X., Pan, J., Zhang, W., Zheng, W., Wang, H. and Gu, Y. (2015) 'Quinolone resistance-determining region mutations and the plasmid-mediated quinolone resistance gene qnrS played important roles in decreased susceptibility to fluoroquinolones among Shigella isolates in southeast China between 1998 and 2013', *International Journal of Antimicrobial Agents*. Elsevier B.V., pp. 2014–2015. doi: 10.1016/j.ijantimicag.2014.12.004.
- Public Health England (2015) 'UK Standards for Microbiology Investigations Identification of Shigella species', pp. 1–22. doi: ID 7.
- Pupo, G. M., Lan, R. and Reeves, P. R. (2000) 'Multiple independent origins of Shigella clones of Escherichia coli and convergent evolution of many of their characteristics', *Proceedings of the National Academy of Sciences of the United States of America*. 2000/08/24, 97(19), pp. 10567–10572. doi: 10.1073/pnas.180094797.
- Qin, T., Qian, H., Fan, W., Ma, P., Zhou, L., Dong, C., Gu, B. and Huo, X. (2017)

'Newest data on fluoroquinolone resistance mechanism of Shigella flexneri isolates in Jiangsu Province of China', *Antimicrobial Resistance and Infection Control*. Antimicrobial Resistance & Infection Control, 6(97), pp. 1–8. doi: 10.1186/s13756-017-0249-1.

- Qiu, S., Wang, Z., Chen, C., Liu, N., Jia, L., Liu, W., Wang, L., Hao, R., Zhang, L., Wang, Y. and Song, H. (2011) 'Emergence of a novel Shigella flexneri serotype 4s strain that evolved from a serotype X variant in China', *Journal of clinical microbiology*. 2010/12/24, 49(3), pp. 1148–1150. doi: 10.1128/JCM.01946-10.
- R Core Team (2016) 'R: A language and environment for statistical computing.' Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.r-project.org/.
- Rahman, M. Z., Azmuda, N., Hossain, M. J., Sultana, M., Khan, S. I. and Birkeland, N. K. (2011) 'Recovery and characterization of environmental variants of Shigella flexneri from surface water in Bangladesh', *Current Microbiology*, 63(4), pp. 372–376. doi: 10.1007/s00284-011-9992-3.
- Rahube, T. O., Viana, L. S., Koraimann, G. and Yost, C. K. (2014) 'Characterization and comparative analysis of antibiotic resistance plasmids isolated from a wastewater treatment plant', *Frontiers in Microbiology*, 5(OCT), pp. 1–9. doi: 10.3389/fmicb.2014.00558.
- Rambaut, A., Lam, T. T., Max Carvalho, L. and Pybus, O. G. (2016) 'Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen)', *Virus Evolution*, 2(1), p. vew007. doi: 10.1093/ve/vew007.
- Ramos, H. C., Rumbo, M. and Sirard, J. C. (2004) 'Bacterial flagellins: Mediators of pathogenicity and host immune responses in mucosa', *Trends in Microbiology*, 12(11), pp. 509–517. doi: 10.1016/j.tim.2004.09.002.
- Ratnayake, R., Allard, R. and Pilon, P. a (2012) 'Shifting dominance of Shigella species in men who have sex with men.', *Epidemiology and infection*, 140(11), pp. 2082–6. doi: 10.1017/S0950268812000738.
- Ravel, J., Gajer, P., Abdo, Z., Schneider, G. M., Koenig, S. S. K., McCulle, S. L., Karlebach, S., Gorle, R., Russell, J., Tacket, C. O., Brotman, R. M., Davis, C. C., Ault, K., Peralta, L. and Forney, L. J. (2011) 'Vaginal microbiome of reproductive-

age women', *Proceedings of the National Academy of Sciences*, 108(Supplement_1), pp. 4680–4687. doi: 10.1073/pnas.1002611107.

- Rehel, N. and Szatmari, G. (1996) 'Characterization of the stable maintenance of the Shigella flexneri plasmid pHS-2', *Plasmid*, 36(36), p. 183–90+. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9007013.
- Revell, L. J. (2012) 'phytools: an R package for phylogenetic comparative biology (and other things)', *Methods in Ecology and Evolution*, 3, pp. 217–223. doi: 10.1111/j.2041-210X.2011.00169.x.
- Rezwan, F., Lan, R. and Reeves, P. R. (2004) 'Molecular basis of the indole-negative reaction in Shigella strains: Extensive damages to the tna operon by insertion sequences', *Journal of Bacteriology*, 186(21), pp. 7460–7465. doi: 10.1128/JB.186.21.7460-7465.2004.
- Ribot, E. M., Fair, M. A., Gautom, R., Cameron, D. N., Hunter, S. B., Swaminathan, B. and Barrett, T. J. (2006) 'Standardization of pulsed-field gel electrophoresis protocols for the subtyping of Escherichia coli O157:H7, Salmonella, and Shigella for PulseNet', *Foodborne pathogens and disease*. 2006/04/11, 3(1), pp. 59–67. doi: 10.1089/fpd.2006.3.59.
- Riddle, M. S., Kaminski, R. W., Williams, C., Porter, C., Baqar, S., Kordis, A., Gilliland, T., Lapa, J., Coughlin, M., Soltis, C., Jones, E., Saunders, J., Keiser, P. B., Ranallo, R. T., Gormley, R., Nelson, M., Turbyfill, K. R., Tribble, D. and Oaks, E. V (2011) 'Safety and immunogenicity of an intranasal Shigella flexneri 2a Invaplex 50 vaccine.', *Vaccine*. Elsevier Ltd, 29(40), pp. 7009–19. doi: 10.1016/j.vaccine.2011.07.033.
- Robrish, S. A., Oliver, C. and Thompson, J. (1991) 'Sugar metabolism by fusobacteria: regulation of transport, phosphorylation, and polymer formation by Fusobacterium mortiferum ATCC 25557.', *Infection and immunity*, 59(12), pp. 4547–54. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=259076&tool=pmcentrez&rendertype=abstract.
- Rohmer, L., Jacobs, M. a, Brittnacher, M. J., Fong, C., Hayden, H. S., Hocquet, D.,

Weiss, E. J., Radey, M., Germani, Y., Talukder, K. A., Hager, A. J., Kemner, J. M., Sims-Day, E. H., Matamouros, S., Hager, K. R. and Miller, S. I. (2014) 'Genomic analysis of the emergence of 20th century epidemic dysentery.', *BMC genomics*, 15, p. 355. doi: 10.1186/1471-2164-15-355.

- Ruekit, S., Wangchuk, S., Dorji, T., Tshering, K. P., Pootong, P., Nobthai, P., Serichantalergs, O., Poramathikul, K., Bodhidatta, L. and Mason, C. J. (2014) 'Molecular characterization and PCR-based replicon typing of multidrug resistant Shigella sonnei isolates from an outbreak in Thimphu , Bhutan', *BMC Research Notes*. BMC Research Notes, 7(1), pp. 1–9. doi: 10.1186/1756-0500-7-95.
- Ruiz, J. (2003) 'Mechanisms of resistance to quinolones: Target alterations, decreased accumulation and DNA gyrase protection', *Journal of Antimicrobial Chemotherapy*, 51(5), pp. 1109–1117. doi: 10.1093/jac/dkg222.
- Sack, D., Hogue, A., Hug, A. and Etheridge, M. (1994) 'Is protection against shigellosis induced by natural infection with Plesiomonas shigelloides', *Lancet*, 343(8910), pp. 1413–5.
- Sadouki, Z., Day, M. R., Doumith, M., Chattaway, M. A., Dallman, T. J., Hopkins, K. L., Elson, R., Woodford, N., Godbole, G. and Jenkins, C. (2017) 'Comparison of phenotypic and WGS-derived antimicrobial resistance profiles of Shigella sonnei isolated from cases of diarrhoeal disease in England and Wales, 2015', *Journal of Antimicrobial Chemotherapy*. doi: 10.1093/jac/dkx170.
- Saeed, A., Abd, H., Edvinsson, B. and Sandstrom, G. (2009) 'Acanthamoeba castellanii an environmental host for Shigella dysenteriae and Shigella sonnei', *Archives of microbiology*. 2008/08/21, 191(1), pp. 83–88. doi: 10.1007/s00203-008-0422-2.
- Sahl, J. W., Morris, C. R., Emberger, J., Fraser, C. M., Ochieng, J. B., Juma, J., Fields, B., Breiman, R. F., Gilmour, M., Nataro, J. P. and Rasko, D. a. (2015) 'Defining the phylogenomics of Shigella species: a pathway to diagnostics.', *Journal of clinical microbiology*, 53(3), pp. 951–60. doi: 10.1128/JCM.03527-14.
- Sansonetti, P. J., Arondel, J., Cantey, J. R., Huerre, M., Pre, M. and Carolina, S. (1996) 'Infection of Rabbit Peyer's Patches by Shigella flexneri : Effect of Adhesive or Invasive Bacterial Phenotypes on Follicle-Associated Epithelium', *Infection and*

immunity, 64(7), pp. 2752–2764.

- Sarker, S. A., Sultana, S., Reuteler, G., Moine, D., Descombes, P., Charton, F., Bourdin, G., McCallin, S., Ngom-Bru, C., Neville, T., Akter, M., Huq, S., Qadri, F., Talukdar, K., Kassam, M., Delley, M., Loiseau, C., Deng, Y., El Aidy, S., Berger, B. and Brüssow, H. (2016) 'Oral Phage Therapy of Acute Bacterial Diarrhea With Two Coliphage Preparations: A Randomized Trial in Children From Bangladesh', *EBioMedicine*, 4, pp. 124–137. doi: 10.1016/j.ebiom.2015.12.023.
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. a, Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J. and Weber, C. F. (2009) 'Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities.', *Applied and environmental microbiology*, 75(23), pp. 7537–41. doi: 10.1128/AEM.01541-09.
- Schroeder, G. N. and Hilbi, H. (2008) 'Molecular pathogenesis of Shigella spp.: controlling host cell signaling, invasion, and death by type III secretion', *Clinical microbiology reviews*. 2008/01/19, 21(1), pp. 134–156. doi: 10.1128/CMR.00032-07.
- Seemann, T. (2014) 'Prokka: rapid prokaryotic genome annotation.', *Bioinformatics (Oxford, England)*, 30(14), pp. 2068–9. doi: 10.1093/bioinformatics/btu153.
- von Seidlein, L., Kim, D. R., Ali, M., Lee, H., Wang, X., Thiem, V. D., Canh, D. G., Chaicumpa, W., Agtini, M. D., Hossain, A., Bhutta, Z. A., Mason, C., Sethabutr, O., Talukder, K., Nair, G. B., Deen, J. L., Kotloff, K., Clemens, J. and Canh do, G. (2006) 'A multicentre study of Shigella diarrhoea in six Asian countries: disease burden, clinical manifestations, and microbiology', *PLoS medicine*. 2006/09/14, 3(9), p. e353. doi: 10.1371/journal.pmed.0030353.
- Sekirov, I., Russell, S. and Antunes, L. (2010) 'Gut microbiota in health and disease', *Physiological Reviews*, 90(3), pp. 859–904. doi: 10.1152/physrev.00045.2009.
- Sethabutr, O., Venkatesan, M., Murphy, G. S., Eampokalap, B., Hoge, C. W. and Echeverria, P. (1993) 'Detection of Shigellae and Enteroinvasive Escherichia coli by Amplification of the Invasion Plasmid Antigen H DNA Sequence in Patients with

Dysentery', *Journal of Infectious Diseases*, 167(2), pp. 458–461. doi: 10.1093/infdis/167.2.458.

- Shepherd, J. G., Wang, L. and Reeves, P. R. (2000) 'Comparison of O-antigen gene clusters of Escherichia coli (Shigella) sonnei and Plesiomonas shigelloides O17: sonnei gained its current plasmid-borne O-antigen genes from P. shigelloides in a recent event', *Infection and immunity*. 2000/09/19, 68(10), pp. 6056–6061. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10992522.
- Shiga, K. (1898) 'Ueber den erreger der dysenterie in Japan (vorläufige mitteilung).', *Zentralbl. Bakteriol. Mikrobiol.*
- Siguier, P., Gourbeyre, E. and Chandler, M. (2014) 'Bacterial insertion sequences: Their genomic impact and diversity', *FEMS Microbiology Reviews*, 38, pp. 865–891. doi: 10.1111/1574-6976.12067.
- Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G. and Ng, P. C. (2012) 'SIFT web server: predicting effects of amino acid substitutions on proteins.', *Nucleic acids research*, 40(Web Server issue), pp. W452-7. doi: 10.1093/nar/gks539.
- Sims, G. E. and Kim, S. H. (2011) 'Whole-genome phylogeny of Escherichia coli/Shigella group by feature frequency profiles (FFPs)', *Proceedings of the National Academy of Sciences of the United States of America*. 2011/05/04, 108(20), pp. 8329–8334. doi: 10.1073/pnas.1105168108.
- Singh, P., Teal, T. K., Marsh, T. L., Tiedje, J. M., Mosci, R., Jernigan, K., Zell, A., Newton, D. W., Salimnia, H., Lephart, P., Sundin, D., Khalife, W., Britton, R. A., Rudrik, J. T. and Manning, S. D. (2015) 'Intestinal microbial communities associated with acute enteric infections and disease recovery.', *Microbiome*. Microbiome, 3(1), p. 45. doi: 10.1186/s40168-015-0109-2.
- Sköld, O. (2011) *Antibiotics and Antibiotic Resistance*. Hoboken, New Jersey: Wiley.
- Skoudy, A., Mounier, J., Aruffo, A., Ohayon, H., Gounon, P., Sansonetti, P. and Tran Van Nhieu, G. (2000) 'CD44 binds to the Shigella IpaB protein and participates in bacterial invasion of epithelial cells', *Cellular Microbiology*, 2(1), pp. 19–33. doi: 10.1046/j.1462-5822.2000.00028.x.
- Smith, A. M., Keddy, K. H., Sooka, A., Ismail, H., De, G. M. and Africa, S. (2009)

'Analysis of a temporal cluster of Shigella boydii isolates in Mpumalanga , South Africa , November to December 2007', *J Infect Dev Ctries*, 3(1), pp. 65–70.

- Stagg, R. M., Cam, P. D. and Verma, N. K. (2008) 'Identification of newly recognized serotype 1c as the most prevalent Shigella flexneri serotype in northern rural Vietnam', *Epidemiology and infection*. 2007/10/10, 136(8), pp. 1134–1140. doi: 10.1017/S0950268807009600.
- Stagg, R. M., Tang, S.-S. S., Carlin, N. I. a, Talukder, K. A., Cam, P. D. and Verma, N. K. (2009) 'A novel glucosyltransferase involved in O-antigen modification of Shigella flexneri serotype 1c.', *Journal of bacteriology*. 2009/09/01, 191(21), pp. 6612–7. doi: 10.1128/JB.00628-09.
- Stamatakis, A. (2014) 'RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.', *Bioinformatics (Oxford, England)*, 30(9), pp. 1312–3. doi: 10.1093/bioinformatics/btu033.
- Sturgill, G., Toutain, C. M., Komperda, J., Toole, G. A. O. and Rather, P. N. (2004) 'Role of CysE in Production of an Extracellular Signaling Molecule in Providencia stuartii and Escherichia coli : Loss of cysE Enhances Biofilm Formation in Escherichia coli', *Journal of Bacteriology*, 186(22), pp. 7610–7617. doi: 10.1128/JB.186.22.7610.
- Subramanian, S., Huq, S., Yatsunenko, T., Haque, R., Mahfuz, M., Alam, M. A., Benezra, A., DeStefano, J., Meier, M. F., Muegge, B. D., Barratt, M. J., VanArendonk, L. G., Zhang, Q., Province, M. A., Petri Jr, W. A., Ahmed, T. and Gordon, J. I. (2014) 'Persistent gut microbiota immaturity in malnourished Bangladeshi children', *Nature*. Nature Publishing Group, 510(7505), pp. 417–421. doi: 10.1038/nature13421.
- Sun, Q., Lan, R., Wang, J., Xia, S., Wang, Y., Wang, Y., Jin, D., Yu, B., Knirel, Y. a and Xu, J. (2013) 'Identification and characterization of a novel Shigella flexneri serotype Yv in China.', *PloS one*, 8(7), p. e70238. doi: 10.1371/journal.pone.0070238.
- Sun, Q., Lan, R., Wang, Y., Wang, J., Li, P., Du, P. and Xu, J. (2013) 'Isolation and genomic characterization of SfI, a serotype-converting bacteriophage of Shigella flexneri', *BMC microbiology*. 2013/02/19, 13(1), p. 39. doi: 10.1186/1471-2180-13-

39.

- Sun, Q., Lan, R., Wang, Y., Zhao, A., Zhang, S., Wang, J., Wang, Y., Xia, S., Jin, D., Cui, Z., Zhao, H., Li, Z., Ye, C., Zhang, S., Jing, H. and Xu, J. (2011) 'Development of a multiplex PCR assay targeting O-antigen modification genes for molecular serotyping of Shigella flexneri', *Journal of Clinical Microbiology*, 49(11), pp. 3766–3770. doi: 10.1128/JCM.01259-11.
- Tacconelli, E., Magrini, N., Kahlmeter, G. and Singh, N. (2017) 'Global Priority List Of Antibiotic-Resistant Bacteria To Guide Research, Discovery, And Development Of New Antibiotics', *World Health Organization*, pp. 1–7.
- Tajbakhsh, M., García Migura, L., Rahbar, M., Svendsen, C. A., Mohammadzadeh, M., Zali, M. R., Aarestrup, F. M. and Hendriksen, R. S. (2012) 'Antimicrobial-resistant Shigella infections from Iran: An overlooked problem?', *Journal of Antimicrobial Chemotherapy*, 67(5), pp. 1128–1133. doi: 10.1093/jac/dks023.
- Talukder, K. a., Islam, Z., Dutta, D. K., Aminul Islam, M., Khajanchi, B. K., Azmi, I. J., Iqbal, M. S., Hossain, M. a., Faruque, a. S. G., Balakrish Nair, G. and Sack, D. a. (2006) 'Antibiotic resistance and genetic diversity of Shigella sonnei isolated from patients with diarrhoea between 1999 and 2003 in Bangladesh', *Journal of Medical Microbiology*, 55(9), pp. 1257–1263. doi: 10.1099/jmm.0.46641-0.
- Taneja, N. (2007) 'Changing Epidemiology of Shigellosis and Emergence of Ciprofloxacin-Resistant Shigellae in India', *Journal of Clinical Microbiology*, 45(2), pp. 678–679. doi: 10.1128/JCM.02247-06.
- Taneja, N. and Mewara, A. (2016) 'Shigellosis : Epidemiology in India', *Indian Journal of Medical Research*, 143(May), pp. 565–576. doi: 10.4103/0971-5916.187104.
- Tang, S.-S., Carlin, N. I. A., Talukder, K. A., Cam, P. D. and Verma, N. K. (2016) 'Shigella flexneri serotype 1c derived from serotype 1a by acquisition of gtrIC gene cluster via a bacteriophage', *BMC Microbiology*. BMC Microbiology, 16(1), p. 127. doi: 10.1186/s12866-016-0746-z.
- Tavaré, S. (1986) 'Some probabilistic and statistical problems in the analysis of DNA sequences', *American Mathematical Society: Lectures on Mathematics in the Life Sciences*, pp. 57–86. doi: citeulike-article-id:4801403.

- Tenzen, T. and Ohtsubo, E. (1991) 'Preferential transposition of an IS630-associated composite transposon to TA in the 5'-CTAG-3' sequence', *Journal of Bacteriology*, 173(19), pp. 6207–6212.
- Thanh, V. T. (2005) 'Vietnam ' s Trade Liberalization and International Economic Integration: Evolution, Problems, and Challenges', *ASEAN Economic Bulletin*, 22(1), pp. 75–91.
- The Human Microbiome Project Consortium (2012) 'Structure, function and diversity of the healthy human microbiome.', *Nature*, 486(7402), pp. 207–14. doi: 10.1038/nature11234.
- Thompson, C. N., Duy, P. T. and Baker, S. (2015) 'The Rising Dominance of Shigella sonnei: An Intercontinental Shift in the Etiology of Bacillary Dysentery', *PLOS Neglected Tropical Diseases*, 9(6), pp. 1–13. doi: 10.1371/journal.pntd.0003708.
- Thompson, C. N., Phan, M. V. T., Hoang, N. V. M., Minh, P. V., Vinh, N. T., Thuy, C. T., Nga, T. T. T., Rabaa, M. a., Duy, P. T., Dung, T. T. N., Phat, V. V., Nga, T. V. T., Tu, L. T. P., Tuyen, H. T., Yoshihara, K., Jenkins, C., Duong, V. T., Phuc, H. L., Tuyet, P. T. N., Ngoc, N. M., Vinh, H., Chinh, N. T., Thuong, T. C., Tuan, H. M., Hien, T. T., Campbell, J. I., Chau, N. V. V., Thwaites, G. and Baker, S. (2015) 'A Prospective Multi-Center Observational Study of Children Hospitalized with Diarrhea in Ho Chi Minh City, Vietnam', *Am. J. Trop. Med. Hyg.*, 92(5), pp. 1045–1052. doi: 10.4269/ajtmh.14-0655.
- Thorsen, J., Brejnrod, A., Mortensen, M., Rasmussen, M. A., Stokholm, J., Al-Soud, W. A., Sørensen, S., Bisgaard, H. and Waage, J. (2016) 'Large-scale benchmarking reveals false discoveries and count transformation sensitivity in 16S rRNA gene amplicon data analysis methods used in microbiome studies', *Microbiome*. Microbiome, 4(1), p. 62. doi: 10.1186/s40168-016-0208-8.
- Tibshirani, R., Walther, G. and Hastie, T. (2001) 'Estimating the number of clusters in a data set via the gap statistic', *Journal of the Royal Statistical Society*, 63(2), pp. 411–423. doi: 10.1111/1467-9868.00293.
- Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., Calteau, A., Chiapello, H., Clermont, O.,

Cruveiller, S., Danchin, A., Diard, M., Dossat, C., Karoui, M. El, Frapy, E., Garry, L., Ghigo, J. M., Gilles, A. M., Johnson, J., Le Bouguenec, C., Lescat, M., Mangenot, S., Martinez-Jehanne, V., Matic, I., Nassif, X., Oztas, S., Petit, M. A., Pichon, C., Rouy, Z., Ruf, C. Saint, Schneider, D., Tourret, J., Vacherie, B., Vallenet, D., Medigue, C., Rocha, E. P. C., Denamur, E., Le Bouguénec, C., Martinez-Jéhanne, V. and Médigue, C. (2009) 'Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths', *PLoS genetics*. 2009/01/24, 5(1), p. e1000344. doi: 10.1371/journal.pgen.1000344.

- Traa, B. S., Fischer Walker, C. L., Munos, M. and Black, R. E. (2010) 'Antibiotics for the treatment of dysentery in children', *International Journal of Epidemiology*, 39(SUPPL. 1), pp. 70–74. doi: 10.1093/ije/dyq024.
- Troeger, C., Forouzanfar, M., Rao, P. C., Khalil, I., Brown, A., Reiner Jr, R. C., Fullman, N., Thompson, R. L., Abajobir, A., Ahmed, M., Alemayohu, M. A., Alvis-Guzman, N., Amare, A. T., Antonio, C. A., Asayesh, H., Avokpaho, E., Awasthi, A., Bacha, U., Barac, A., Betsue, B. D., Beyene, A. S., Boneya, D. J., Malta, D. C., Dandona, L., Dandona, R., Dubey, M., Eshrati, B., Fitchett, J. R. A., Gebrehiwot, T. T., Hailu, G. B., Horino, M., Hotez, P. J., Jibat, T., Jonas, J. B., Kasaeian, A., Kissoon, N., Kotloff, K., Koyanagi, A., Kumar, G. A., Rai, R. K., Lal, A., El Razek, H. M. A., Mengistie, M. A., Moe, C., Patton, G., Platts-Mills, J. A., Qorbani, M., Ram, U., Roba, H. S., Sanabria, J., Sartorius, B., Sawhney, M., Shigematsu, M., Sreeramareddy, C., Swaminathan, S., Tedla, B. A., Jagiellonian, R. T.-M., Ukwaja, K., Werdecker, A., Widdowson, M.-A., Yonemoto, N., El Sayed Zaki, M., Lim, S. S., Naghavi, M., Vos, T., Hay, S. I., Murray, C. J. L. and Mokdad, A. H. (2017) 'Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015', *The Lancet Infectious Diseases*, 3099(17), pp. 1–40. doi: https://doi.org/10.1016/S1473-3099(17)30276-1.
- Trofa, A. F., Ueno-olsen, H., Oiwa, R. and Yoshikawa, M. (1999) 'Dr . Kiyoshi Shiga : Discoverer of the Dysentery Bacillus', *Clincal Infectious Diseases*, 29, pp. 1303–1306.
- Turnbaugh, P. J., Ley, R. E., Mahowald, M. a, Magrini, V., Mardis, E. R. and

Gordon, J. I. (2006) ‘An obesity-associated gut microbiome with increased capacity for energy harvest.’, *Nature*, 444(7122), pp. 1027–31. doi: 10.1038/nature05414.

- Uchimura, M., Kishida, K. and Koiwai, K. (2001) ‘Increasing Incidence and the Mechanism of Resistance of Nalidixic Acid Resistant Shigella sonnei’, *Kansenshogaku Zasshi*, 75, pp. 923–930.
- Ud-Din, A. I. M. S., Wahid, S. U. H., Latif, H. a, Shahnaij, M., Akter, M., Azmi, I. J., Hasan, T. N., Ahmed, D., Hossain, M. a, Faruque, A. S. G., Faruque, S. M. and Talukder, K. a (2013) ‘Changing trends in the prevalence of Shigella species: emergence of multi-drug resistant Shigella sonnei biotype g in Bangladesh.’, *PloS one*, 8(12), p. e82601. doi: 10.1371/journal.pone.0082601.
- Velly, H., Britton, R. A. and Preidis, G. A. (2017) ‘Mechanisms of cross-talk between the diet , the intestinal microbiome , and the undernourished host’, *Gut Microbes*. Taylor & Francis, 8(2), pp. 98–112. doi: 10.1080/19490976.2016.1267888.
- Venables, W. . and Ripley, B. . (2002) *Modern Applied Statistics with S*. Fourth Edi. New York: Springer.
- Venkatesan, M. M., Goldberg, M. B., Rose, D. J., Grotbeck, E. J., Burland, V. and Blattner, F. R. (2001) ‘Complete DNA Sequence and Analysis of the Large Virulence Plasmid of Shigella flexneri Complete DNA Sequence and Analysis of the Large Virulence Plasmid of Shigella flexneri’, *Infection and Immunity*. doi: 10.1128/IAI.69.5.3271.
- Vinh, H., Anh, V. T. C., Anh, N. D., Campbell, J. I., van Minh Hoang, N., Nga, T. V. T., Nhu, N. T. K., van Minh, P., Thuy, C. T., Duy, P. T., Phuong, L. T., Loan, H. T., Chinh, M. T., Thao, N. T. T., Tham, N. T. H., Mong, B. L., van Be Bay, P., Day, J. N., Dolecek, C., Lan, N. P. H., Diep, T. S., Farrar, J. J., van Vinh Chau, N., Wolbers, M. and Baker, S. (2011) ‘A multi-center randomized trial to assess the efficacy of gatifloxacin versus ciprofloxacin for the treatment of shigellosis in Vietnamese children’, *PLoS Neglected Tropical Diseases*, 5(8). doi: 10.1371/journal.pntd.0001264.
- Vinh, H., Main, J., Chinh, M., Tam, C., Trang, P., Nga, D., Echeverria, P., Diep, T., White, N. and Parry, C. (2000) ‘Treatment of bacillary dysentery in Vietnamese

children: two doses of ofloxacin versus 5-days nalidixic acid', *Trans R Soc Trop Med Hyg*, 94(3), pp. 323–326.

- Vinh, H., Nhu, N. T., Nga, T. V, Duy, P. T., Campbell, J. I., Hoang, N. V, Boni, M. F., My, P. V, Parry, C., Nga, T. T., Van Minh, P., Thuy, C. T., Diep, T. S., Phuong le, T., Chinh, M. T., Loan, H. T., Tham, N. T., Lanh, M. N., Mong, B. L., Anh, V. T., Bay, P. V, Chau, N. V, Farrar, J. and Baker, S. (2009) 'A changing picture of shigellosis in southern Vietnam: shifting species dominance, antimicrobial susceptibility and clinical presentation', *BMC infectious diseases*. 2009/12/17, 9, p. 204. doi: 10.1186/1471-2334-9-204.
- Vokes, S. A., Reeves, S. A., Torres, A. G. and Payne, S. M. (1999) 'The aerobactin iron transport system genes in Shigella flexneri are present within a pathogenicity island', *Molecular microbiology*. 1999/07/20, 33(1), pp. 63–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10411724.
- Vrints, M., Mairiaux, E., Van Meervenne, E., Collard, J. M. and Bertrand, S. (2009) 'Surveillance of antibiotic susceptibility patterns among Shigella sonnei strains isolated in Belgium during the 18-year period 1990 to 2007', *Journal of Clinical Microbiology*, 47(5), pp. 1379–1385. doi: 10.1128/JCM.02460-08.
- Vu, D. T., Sethabutr, O., von Seidlein, L., Tran, V. T., Do, G. C., Bui, T. C., Le, H. T., Lee, H., Mason, C. and Dang, D. T. (2004) 'Detection of Shigella by a PCR Assay Targeting the ipaH Gene Suggests Increased Prevalence of Shigellosis in Nha Trang, Vietnam', *Journal of clinical microbiology*, 42(5), pp. 2031–2035. doi: 10.1128/JCM.42.5.2031.
- Vuong, H. E. and Hsiao, E. Y. (2017) 'Emerging Roles for the Gut Microbiome in Autism Spectrum Disorder', *Biological Psychiatry*. Elsevier, 81(5), pp. 411–423. doi: 10.1016/j.biopsych.2016.08.024.
- Wagner, A. and de la Chaux, N. (2008) 'Distant horizontal gene transfer is rare for multiple families of prokaryotic insertion sequences', *Molecular genetics and genomics : MGG*. 2008/08/30, 280(5), pp. 397–408. doi: 10.1007/s00438-008-0373-y.
- Wang, Q., Garrity, G. M., Tiedje, J. M. and Cole, J. R. (2007) 'Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy.',

Applied and environmental microbiology, 73(16), pp. 5261–7. doi: 10.1128/AEM.00062-07.

- Wang, X. Y., Du, L., Von Seidlein, L., Xu, Z. Y., Zhang, Y. L., Hao, Z. Y., Han, O. P., Ma, J. C., Lee, H. J., Ali, M., Han, C. Q., Xing, Z. C., Chen, J. C. and Clemens, J. (2005) 'Occurrence of shigellosis in the young and elderly in rural China: Results of a 12-month population-based surveillance study', *American Journal of Tropical Medicine and Hygiene*, 73(2), pp. 416–422.
- Watarai, M., Funato, S. and Sasakawa, C. (1996) 'Interaction of Ipa proteins of Shigella flexneri with α5β1 integrin promotes entry of the bacteria into mammalian cells.', *The Journal of experimenta Medicine*, 183(March), pp. 991–999. Available at: http://onlinelibrary.wiley.com/doi/10.1093/emboj/cdf457/full (Accessed: 9 January 2015).
- Wehler, T. and Carlin, N. I. (1988) 'Structural and immunochemical studies of the lipopolysaccharide from a new provisional serotype of Shigella flexneri', *European journal of biochemistry / FEBS*. 1988/09/15, 176(2), pp. 471–476. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2458260.
- Wei, H. L., Wang, Y. W., Li, C. C., Tung, S. K. and Chiou, C. S. (2007) 'Epidemiology and evolution of genotype and antimicrobial resistance of an imported Shigella sonnei clone circulating in central Taiwan', *Diagnostic Microbiology and Infectious Disease*, 58, pp. 469–475. doi: 10.1016/j.diagmicrobio.2007.03.006.
- Weill, F., Domman, D., Njamkepo, E., Tarr, C., Rauzier, J., Fawal, N., Keddy, K. H., Salje, H., Moore, S., Mukhopadhyay, A. K., Bercion, R., Luquero, F. J., Koeck, J., Fournier, J. and Dougan, G. (2017) 'Genomic history of the seventh pandemic of cholera in Africa', *Science*, 789(November), pp. 785–789.
- Weiss, S., Treuren, W. Van, Lozupone, C., Faust, K., Friedman, J., Deng, Y., Xia, L. C., Xu, Z. Z. and Ursell, L. (2016) 'Correlation detection strategies in microbial data sets vary widely in sensitivity and precision', *The ISME Journal*. Nature Publishing Group, 10(7), pp. 1669–1681. doi: 10.1038/ismej.2015.235.
- Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J. R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E. R. and Knight, R.

(2017) ‘Normalization and microbial differential abundance strategies depend upon data characteristics’, *Microbiome*. Microbiome, 5(1), p. 27. doi: 10.1186/s40168-017-0237-y.

- West, N. P., Sansonetti, P., Mounier, J., Exley, R. M., Parsot, C., Guadagnini, S., Prévost, M.-C., Prochnicka-Chalufour, A., Delepierre, M., Tanguy, M. and Tang, C. M. (2005) ‘Optimization of virulence functions through glucosylation of Shigella LPS.’, *Science*, 307(5713), pp. 1313–1317. doi: 10.1126/science.1108472.
- Whiteway, J., Koziarz, P., Veall, J., Sandhu, N., Kumar, P., Hoecher, B. and Lambert, I. B. (1998) ‘Oxygen-insensitive nitroreductases: Analysis of the roles of nfsA and nfsB in development of resistance to 5-nitrofuran derivatives in Escherichia coli’, *Journal of Bacteriology*, 180(21), pp. 5529–5539.
- WHO (2005) *Guidelines for the control of shigellosis, including epidemics due to Shigella dysenteriae type 1*, *World Health Organization*.
- WHO (2016) ‘Global Health Estimates 2015: Disease burden by Cause, Age, Sex, by Country and by Region, 2000-2015.’ Geneva, Switzerland.
- Wickham, H. (2009) *ggplot2*, *Elegant Graphics for Data Analysis*. doi: 10.1007/978-0-387-98141-3.
- Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L. H., Karch, H., Reeves, P. R., Maiden, M. C., Ochman, H. and Achtman, M. (2006) ‘Sex and virulence in Escherichia coli: an evolutionary perspective’, *Molecular microbiology*. 2006/05/13, 60(5), pp. 1136–1151. doi: 10.1111/j.1365-2958.2006.05172.x.
- Wong, V. K., Baker, S., Pickard, D. J., Parkhill, J., Page, A. J., Feasey, N. a, Kingsley, R. a, Thomson, N. R., Keane, J. a, Weill, F.-X., Edwards, D. J., Hawkey, J., Harris, S. R., Mather, A. E., Cain, A. K., Hadfield, J., Hart, P. J., Thieu, N. T. V., Klemm, E. J., Glinos, D. a, Breiman, R. F., Watson, C. H., Kariuki, S., Gordon, M. a, Heyderman, R. S., Okoro, C., Jacobs, J., Lunguya, O., Edmunds, W. J., Msefula, C., Chabalgoity, J. a, Kama, M., Jenkins, K., Dutta, S., Marks, F., Campos, J., Thompson, C., Obaro, S., MacLennan, C. a, Dolecek, C., Keddy, K. H., Smith, A. M., Parry, C. M., Karkey, A., Mulholland, E. K., Campbell, J. I., Dongol, S., Basnyat, B., Dufour, M., Bandaranayake, D., Naseri, T. T., Singh, S. P., Hatta, M., Newton, P., Onsare, R. S., Isaia, L., Dance, D., Davong, V., Thwaites, G., Wijedoru,

L., Crump, J. a, De Pinna, E., Nair, S., Nilles, E. J., Thanh, D. P., Turner, P., Soeng, S., Valcanis, M., Powling, J., Dimovski, K., Hogg, G., Farrar, J., Holt, K. E. and Dougan, G. (2015) 'Phylogeographical analysis of the dominant multidrug-resistant H58 clade of Salmonella Typhi identifies inter- and intracontinental transmission events', *Nature Genetics*. doi: 10.1038/ng.3281.

- Woodward, D. L., Clark, C. G., Caldeira, R. a., Ahmed, R., Soule, G., Bryden, L., Tabor, H., Melito, P., Foster, R., Walsh, J., Ng, L. K., Malcolm, G. B., Strockbine, N. and Rodgers, F. G. (2005) 'Identification and characterization of Shigella boydii 20 serovar nov., a new and emerging Shigella serotype', *Journal of Medical Microbiology*, 54(8), pp. 741–748. doi: 10.1099/jmm.0.46095-0.
- World Bank (2015) *World Development Indicators*. Available at: https://data.worldbank.org/indicator/.
- World Health Organization (2006) *WHO child growth standards: Methods and Development*, *World Health Organization*. Geneva, Switzerland. doi: 10.4067/S0370-41062009000400012.
- Wyckoff, E. E., Duncan, D., Torres, a G., Mills, M., Maase, K. and Payne, S. M. (1998) 'Structure of the Shigella dysenteriae haem transport locus and its phylogenetic distribution in enteric bacteria.', *Molecular microbiology*, 28(6), pp. 1139–52. doi: 10.1046/j.1365-2958.1998.00873.x.
- Yang, F., Yang, J., Zhang, X., Chen, L., Jiang, Y., Yan, Y., Tang, X., Wang, J., Xiong, Z., Dong, J., Xue, Y., Zhu, Y., Xu, X., Sun, L., Chen, S., Nie, H., Peng, J., Xu, J., Wang, Y., Yuan, Z., Wen, Y., Yao, Z., Shen, Y., Qiang, B., Hou, Y., Yu, J. and Jin, Q. (2005) 'Genome dynamics and diversity of Shigella species, the etiologic agents of bacillary dysentery', *Nucleic acids research*. 2005/11/09, 33(19), pp. 6445–6458. doi: 10.1093/nar/gki954.
- Yang, J., Nie, H., Chen, L., Zhang, X., Yang, F., Xu, X., Zhu, Y., Yu, J. and Jin, Q. (2007) 'Revisiting the molecular evolutionary history of Shigella spp', *Journal of molecular evolution*. 2006/12/13, 64(1), pp. 71–79. doi: 10.1007/s00239-006-0052-8.
- Yang, Z. (2007) 'PAML 4: Phylogenetic analysis by maximum likelihood', *Molecular Biology and Evolution*, 24(8), pp. 1586–1591. doi: 10.1093/molbev/msm088.

- Yatsunenko, T., Rey, F. E., Manary, M. J. M. M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., Heath, A. C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J. G., Lozupone, C. a, Lauber, C., Clemente, J. C., Knights, D., Knight, R. and Gordon, J. I. (2012) 'Human gut microbiome viewed across age and geography', *Nature*, 486(Ivic), pp. 222–7. doi: 10.1038/nature11053.
- Ye, C., Lan, R., Xia, S., Zhang, J., Sun, Q., Zhang, S., Jing, H., Wang, L., Li, Z., Zhou, Z., Zhao, A., Cui, Z., Cao, J., Jin, D., Huang, L., Wang, Y. Y. Y., Luo, X., Bai, X., Wang, P., Xu, Q., Xu, J., Wang, Y. Y. Y., Wang, P., Xu, Q. and Xu, J. (2010) 'Emergence of a new multidrug-resistant serotype X variant in an epidemic clone of Shigella flexneri.', *Journal of clinical microbiology*. 2009/12/04, 48(2), pp. 419–26. doi: 10.1128/JCM.00614-09.
- Yoshida, S. and Katayama, E. (2002) 'Shigella deliver an effector protein to trigger host microtubule destabilization, which promotes Rac1 activity and efficient bacterial internalization', *The EMBO Journal*, 21(12). Available at: http://emboj.embopress.org/content/21/12/2923.abstract (Accessed: 9 January 2015).
- Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F. M. and Larsen, M. V. (2012) 'Identification of acquired antimicrobial resistance genes.', *The Journal of antimicrobial chemotherapy*, 67(11), pp. 2640–4. doi: 10.1093/jac/dks261.
- Zeileis, A. and Hothorn, T. (2002) 'Diagnostic Checking in Regression Relationships', *R News*, 2(3), pp. 7–10.
- Zerbino, D. R. and Birney, E. (2008) 'Velvet: algorithms for de novo short read assembly using de Bruijn graphs.', *Genome research*, 18(5), pp. 821–9. doi: 10.1101/gr.074492.107.
- Zhang, N., Lan, R., Sun, Q., Wang, J., Wang, Y., Zhang, J., Yu, D., Hu, W., Hu, S., Dai, H., Du, P., Wang, H. and Xu, J. (2014) 'Genomic portrait of the evolution and epidemic spread of a recently emerged multidrug-resistant Shigella flexneri clone in China.', *Journal of clinical microbiology*, 52(4), pp. 1119–26. doi: 10.1128/JCM.02669-13.
- Zhang, Y. and Lin, K. (2012) 'A phylogenomic analysis of Escherichia coli /

Shigella group: implications of genomic features associated with pathogenicity and ecological adaptation', *BMC evolutionary biology*. 2012/09/11, 12, p. 174. doi: 10.1186/1471-2148-12-174.

- Zhao, G., Zhu, L., Feng, E., Cao, X., Shang, N., Liu, X., Liao, X., Ying, T., Wang, J., Chen, H. and Wang, H. (2010) 'A novel anti-virulence gene revealed by proteomic analysis in Shigella flexneri 2a.', *Proteome science*, 8, p. 30. doi: 10.1186/1477-5956-8-30.
- Zhao, Y., Zhu, Y., Zhu, Z. and Qu, B. (2016) 'Association between meteorological factors and bacillary dysentery incidence in Chaoyang city, China: An ecological study', *BMJ Open*, 6(12). doi: 10.1136/bmjopen-2016-013376.
- Zwickl, D. J. (2006) 'Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion', *Philosophy*, p. 115. doi: (http://www.zo.utexas.edu/faculty/antisense/garli/Garli.html).
- Zychlinsky, A., Prevost, M. and Sansonetti, P. (1992) 'Shigella flexneri induces apoptosis in infected macrophages', *Nature*, 358. Available at: http://www.nature.com/nature/journal/v358/n6382/abs/358167a0.html (Accessed: 9 January 2015).

9 Appendices

9.1 Publications as first authors during DPhil candidature

Published online: February 20, 2015

Research Article

TRANSPARENT PROCESS OPEN ACCESS

EMBO Molecular Medicine

A high-resolution genomic analysis of multidrug-resistant hospital outbreaks of *Klebsiella pneumoniae*

Hao Chung The[1,†], Abhilasha Karkey[2,†], Duy Pham Thanh[1], Christine J Boinett[3], Amy K Cain[3], Matthew Ellington[3,4], Kate S Baker[3], Sabina Dongol[2], Corinne Thompson[1,5], Simon R Harris[3], Thibaut Jombart[6], Tu Le Thi Phuong[1], Nhu Tran Do Hoang[1], Tuyen Ha Thanh[1], Shrijana Shretha[2], Suchita Joshi[2], Buddha Basnyat[2], Guy Thwaites[1,5], Nicholas R Thomson[3,7,‡], Maia A Rabaa[1,8,‡] & Stephen Baker[1,5,7,‡,*]

Abstract

Multidrug-resistant (MDR) *Klebsiella pneumoniae* has become a leading cause of nosocomial infections worldwide. Despite its prominence, little is known about the genetic diversity of *K. pneumoniae* in resource-poor hospital settings. Through whole-genome sequencing (WGS), we reconstructed an outbreak of MDR *K. pneumoniae* occurring on high-dependency wards in a hospital in Kathmandu during 2012 with a case-fatality rate of 75%. The WGS analysis permitted the identification of two MDR *K. pneumoniae* lineages causing distinct outbreaks within the complex endemic *K. pneumoniae*. Using phylogenetic reconstruction and lineage-specific PCR, our data predicted a scenario in which *K. pneumoniae*, circulating for 6 months before the outbreak, underwent a series of ward-specific clonal expansions after the acquisition of genes facilitating virulence and MDR. We suggest that the early detection of a specific NDM-1 containing lineage in 2011 would have alerted the high-dependency ward staff to intervene. We argue that some form of real-time genetic characterisation, alongside clade-specific PCR during an outbreak, should be factored into future healthcare infection control practices in both high- and low-income settings.

Keywords antimicrobial resistance; bloodstream infections; carbapenemases; *Klebsiella pneumoniae*; nosocomial infections

Subject Categories Chromatin, Epigenetics, Genomics & Functional Genomics; Microbiology, Virology & Host Pathogen Interaction

DOI 10.15252/emmm.201404767 | Received 21 October 2014 | Revised 19 January 2015 | Accepted 20 January 2015

Introduction

Klebsiella pneumoniae is one of the leading causes of hospital-acquired infections globally. These infections (typically pneumonia, urinary tract infections, bacteraemia and wound infections) are frequently severe, notoriously affecting incapacitated patients with a suppressed immune status and neonates in intensive care units (Podschun & Ullmann, 1998). *Klebsiella pneumoniae* are commonly resistant to multiple antimicrobials and are generally acknowledged to be a major source of antimicrobial resistance genes that can spread into other Gram-negative pathogens. This has led to multi-drug-resistant (MDR) *K. pneumoniae* being recently singled out as an "urgent threat to human health" by the U.S. Centers for Disease Control and Prevention (CDC), the World Health Organization (WHO), the Government of the United Kingdom and other international healthcare organizations (Centres for Disease Control & Prevention, 2013; UK Department of Health & Department for Environment Food & Rural Affairs, 2013). *Klebsiella pneumoniae* has earned notoriety largely due to resistance to "last resort" antimicrobials such as 3rd generation cephalosporins and carbapenems, which has significantly narrowed, or in some settings completely removed, the therapeutic options for the treatment of MDR *K. pneumoniae* infections (Munoz-Price *et al*, 2013).

1 The Hospital for Tropical Diseases, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam
2 Patan Academy of Health Sciences, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Kathmandu, Nepal
3 The Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK
4 Addenbrooke's Hospital, Cambridge, UK
5 Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, Oxford University, Oxford, UK
6 MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College, London, UK
7 The London School of Hygiene and Tropical Medicine, London, UK
8 Centre for Immunity, Infection and Evolution, University of Edinburgh, Edinburgh, UK
*Corresponding author. Tel: +84 8 39239210; E-mail: sbaker@oucru.org
†These authors contributed equally to this work
‡Joint senior authors

© 2015 The Authors. Published under the terms of the CC BY 4.0 license

The problem of antimicrobial resistance is highlighted by a recent increase of carbapenem-resistant *K. pneumoniae*, which has largely been driven by the emergence and spread of mobile genetic elements carrying carbapenemase resistance genes including the metallo-beta-lactamase NDM-1 (Nordmann *et al*, 2011; Tzouvelekis *et al*, 2012). Indeed, $bla_{\text{NDM-1}}$ is now considered to be endemic in the Indian subcontinent and reports of NDM-1-producing *Klebsiella* from hospital- and community-acquired infections elsewhere indicate global dispersion (Walsh *et al*, 2011; Villa *et al*, 2012; Hudson *et al*, 2014). Perhaps of additional concern is the fact that $bla_{\text{NDM-1}}$ is frequently coupled with several antimicrobial resistance determinants on the same conjugative plasmid, thus presenting a bleak therapeutic outlook for severe infections caused by *K. pneumoniae* carrying such plasmids.

The genetic diversity of *K. pneumoniae* causing hospital outbreaks has been studied previously, with the primary focus being particular emergent clones in affluent settings (Brisse *et al*, 2009; Snitkin *et al*, 2012; Deleo *et al*, 2014). There is currently limited knowledge regarding the overall genetic diversity of this major pathogen within a single hospital setting. This limitation is highly pertinent in resource-poor settings where these infections may exhibit more antimicrobial resistance and cause particularly aggressive infections. This knowledge gap is due in part to a lack of sensitive phenotypic and molecular tools for accurately differentiating or subtyping *K. pneumoniae*. For most Gram-negative bacteria, relatedness among isolates has traditionally been deduced using antimicrobial susceptibility patterns, crude microbiological methods or molecular typing methods such as pulsed field gel electrophoresis (PFGE) or multilocus sequence typing (MLST). Whilst providing some relevant information, these methods rarely provide sufficient discriminatory power for characterising and precisely tracking an outbreak. This is a major constraint that needs to be addressed if we are to make any substantial impact on understanding the genomics and epidemiology of such an important and neglected pathogen, especially if the current trajectory of antimicrobial resistance is sustained.

Here, we have used whole-genome sequencing (WGS) and state-of-the-art analytical techniques to define an alarming outbreak of MDR *K. pneumoniae* within a major hospital in Kathmandu, Nepal, during 2012. This outbreak occurred in children admitted to neighbouring high-dependency paediatric wards and was characterised by high incidence of bloodstream infections and a case-fatality rate of 75% (36/48). At the time, this was hypothesised to be an outbreak caused by a single strain, but we show that it was in fact two outbreaks triggered by two distantly related MDR *K. pneumoniae* lineages against a complex backdrop of sporadic endemic *K. pneumoniae* infections. This study provides essential insights into antimicrobial resistance, virulence, and pre- and post-outbreak strain persistence of an increasingly important global pathogen in a typical limited resource hospital in Kathmandu, Nepal.

Results

Clinical description of *Klebsiella pneumoniae* outbreak

From May 2012, a proportional increase in the number of *Klebsiella* spp. bloodstream infections in children admitted to the several proximal high-dependency wards in Patan Hospital was reported (Fig 1A). The specific wards affected were the nurseries, the neonatal intensive care unit (NICU) and the paediatric intensive care unit (PICU). Indeed, almost 60% (14/24) of the *Klebsiella* spp. isolated in the hospital in May/June 2012 were from blood, compared to an overall average of 20% (81/386) for 2012 (Fig 1B). The bloodstream-associated *Klebsiella* isolates in May/June 2012 were confirmed as *K. pneumoniae* and were first noted to be cefotaxime resistant with variable susceptibility to carbapenems (9/14; 64%). An outbreak was suspected, and as a result, the NICU and the PICU were closed for maintenance (including the repair of leaking pipework) and fumigation (NICU closed, 4/7/12–24/7/12; PICU closed, 9/7/12–24/7/12). Later in 2012, we observed a secondary rise in the number of *K. pneumoniae* isolations in blood taken from children on the same high-dependency wards (50% (10/20) of all *Klebsiella* spp. isolated in the hospital in November 2012) (Fig 1B). At this time, a recrudescence of the earlier outbreak was suspected, given that the *K. pneumoniae* originated from the same wards and were again resistant to cefotaxime.

On review, we found that 270 *Klebsiella* spp. isolates had been cultured across the hospital in 2012, of which 18% (48/270) were *K. pneumoniae* originating from blood cultures taken in the three nurseries, the NICU and the PICU (Fig 1A). Of these 48 cases of childhood *K. pneumoniae* bloodstream infections, 36 (75%) died in hospital. To better understand the nature of this dual peaked *K. pneumoniae* outbreak, we randomly selected 90 hospital *K. pneumoniae* isolates between May and December 2012 for WGS, which included 29 suspected outbreak strains from blood (isolated from blood on the NICU, $n = 8$; PICU, $n = 7$; nurseries, $n = 14$) (isolate information shown in Supplementary Table S1).

Genome sequencing of *Klebsiella pneumoniae* outbreak strains

The 90 selected *K. pneumoniae* strains were subjected to short-read Illumina sequencing. Read mapping to *K. pneumoniae* reference sequence MGH78578 showed an average of 91.12% genome coverage among 89 genomes, while one strain failed due to insufficient coverage. Reads were *de novo* assembled into contigs, and the pan and core genome were constructed by protein clustering (see Materials and Methods). The MLST gene alleles were extracted, and a phylogeny using the seven concatenated MLST gene sequences of all 89 strains in comparison to *K. variicola* (KpIII) and *K. pneumoniae* MGH78578 (KpI) showed that KpI was nested within the 89 hospital strains, while KpIII was clustered as an out-group to all others (unpublished observations). This indicated that all 89 strains isolated were related and also confirmed that they were all restricted to the *K. pneumoniae* species (Brisse *et al*, 2009).

We defined the core (common) genome of the 89 sequenced *K. pneumoniae* as the concatenation of 3,692 gap-free putative orthologous coding sequences, constituting 3,529,263 base pairs of DNA sequence. Since two of the genome sequences showed no evidence of plasmids (reference genome *K. pneumoniae* MGH78578), the core genome contained minimal sequences associated with mobile genetic elements. In this reference-free alignment, we identified 101,250 SNPs across all strains, with the pairwise disparity between strains ranging from 2 to 21,373 SNPs. Putative recombination loci were further detected and removed, and the final post-recombination removal alignment consisted of 83,859 variable SNP sites among the *K. pneumoniae*. This recombination-free SNP variation across the core genome was used to construct a maximum

 © 2015 The Authors

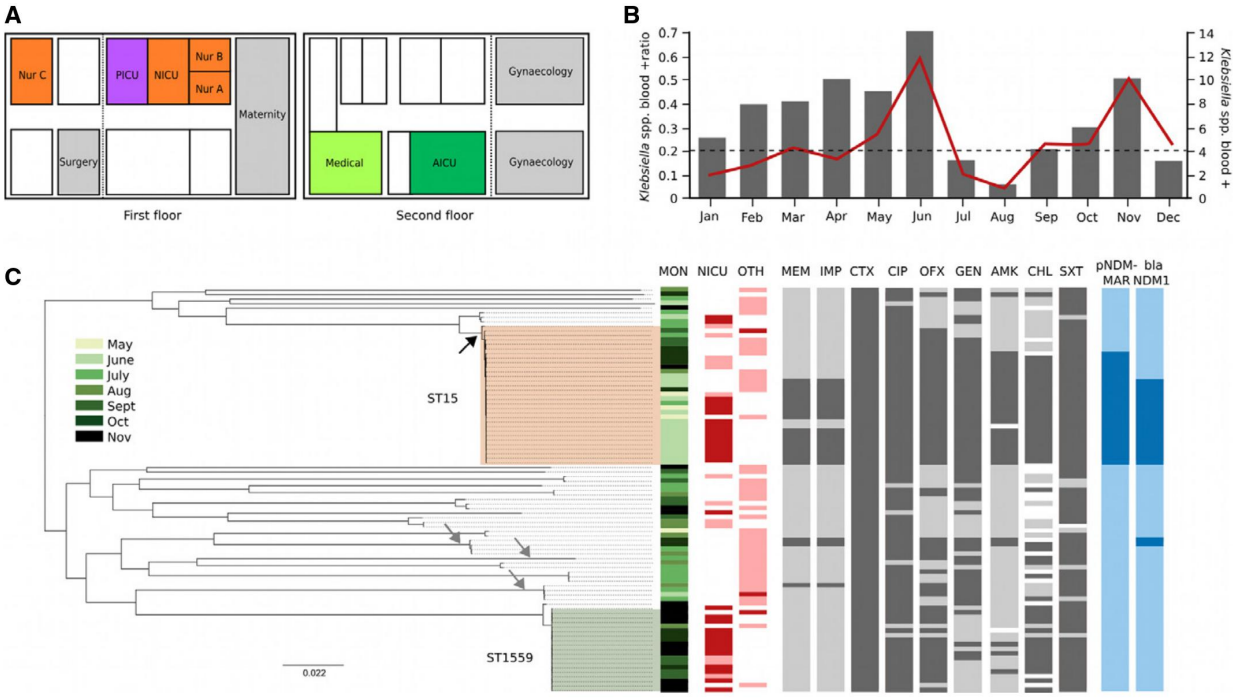


Figure 1. Hospital outbreaks of *Klebsiella pneumoniae* in Kathmandu, Nepal, in 2012.

A Graphical representation of the layout of the wards at Patan Hospital affected by the 2012 *K. pneumoniae* outbreaks. The affected wards on the first and second floor (directly above) are highlighted in colour, which correspond with the sub-lineages indicated in Fig 2A; nurseries and neonatal intensive care unit (NICU), orange; paediatric intensive care unit (PICU), purple; adult intensive care unit (AICU), dark green; medical ward, bright green.

B Plot (red line) shows the monthly ratio of *Klebsiella* spp. isolated from blood at Patan Hospital in 2012 with respect to *Klebsiella* spp. isolated from all other samples (primary *y*-axis). Broken line corresponds to the mean ratio of *Klebsiella* spp. isolated from blood in 2012. Histogram shows the number of *Klebsiella* spp. isolated from blood samples at Patan Hospital per month in 2012 (secondary *y*-axis).

C Maximum likelihood phylogeny based on core genome SNPs of the 89 *K. pneumoniae* isolates with the corresponding metadata: month of isolation (MON) in 2012 (see key), ward (NICU, PICU and nurseries (NICU) or other (OTH)) and specimen type from which the organism was isolated (red, blood; pink, others), antimicrobial susceptibility profile (susceptible, light grey; non-susceptible, dark grey; and missing data, white) by disc diffusion against MEM (meropenem), IMP (imipenem), CTX (cefotaxime), CIP (ciprofloxacin), OFX (ofloxacin), GEN (gentamicin), AMK (amikacin), CHL (chloramphenicol) and SXT (trimethoprim/sulfamethoxazole). The presence (dark blue) or absence (light blue) of the pNDM-MAR-like plasmid and the bla_{NDM-1} gene. The black arrow distinguishes the lineage with three mutations in genes associated with fluoroquinolone susceptibility (GyrA-S83F, GyrA-S87A, ParC-S80I). The grey arrows identify lineages with two mutations in genes associated with fluoroquinolone susceptibility (GyrA-S83I, ParC-S80I). The scale bar indicates the number of nucleotide changes per site (see Materials and Methods).

likelihood phylogenetic tree of all 89 sequenced *K. pneumoniae* (Fig 1C) and revealed a structure dominated by two major genetically distinct lineages.

We next overlaid supplementary metadata onto the whole-genome phylogeny, which included the date of isolation, ward of admission, specimen type and antimicrobial susceptibility against meropenem, imipenem, cefotaxime, ciprofloxacin, ofloxacin, gentamicin, amikacin, chloramphenicol and trimethoprim/sulfamethoxazole (Fig 1C). The first dominant lineage in this *K. pneumoniae* collection belonged to MLST-sequence type (ST)15 ($n = 31$), of which 17 were phenotypically resistant to carbapenems, which appeared to be associated with a pNDM-MAR plasmid containing the bla_{NDM-1} carbapenemase gene. Notably, these ST15 strains were resistant to all other antimicrobials tested (Fig 1C). Fifteen of these ST15 strains were associated with bloodstream infections, and the majority were associated with the NICU complex (NICU, PICU and nurseries) (Fig 1C). ST15 is an individual clonal group (CG) within the MLST clonal complex 14 (CC14) (Bialek-Davenet *et al* 2014), an internationally dispersed clonal complex that has previously been associated with hospital-acquired infections (Hrabák *et al*, 2009).

The second major lineage was found to be a novel sequence type (designated as ST1559), of which all 19 were carbapenem susceptible (variable susceptibility was observed against the other seven antimicrobials tested) (Fig 1C). ST1559 is a single locus variant of ST152, with mutation in the *ton*B-56 allele (T261G). Fourteen of these ST1559 organisms were associated with bloodstream infections and were again mainly associated with the NICU complex (Fig 1C). These data show that the increased incidence of *K. pneumoniae* bloodstream infections on the paediatric wards in Patan Hospital in 2012 was in fact due to two independent outbreaks caused by distinct and widespread lineages.

Defining outbreak lineages

The 31 ST15 isolates predominated and were highly temporally associated; therefore, we constructed a refined maximum

© 2015 The Authors

likelihood phylogeny to study these isolates in greater detail. The phylogeny emphasised a highly structured clonal population composed of 25/31 isolates (14 from blood), which we designated Outbreak Lineage 1. A core genome alignment of these strains revealed a total of 66 non-recombinant SNPs among 24 strains (one strain was excluded due to extensive recombination). Notably, the Outbreak Lineage 1 isolates originating from the adult intensive care unit (AICU) and the PICU fell into distinct clades, as did the NICU and nursery isolates—despite their apparent proximity in time and locality (Fig 2A). These data predict that all of these cases were unlikely to represent an on-going transmission chain of the NICU-based strains outside of the NICU. Nevertheless, an additional strain within the Outbreak Lineage 1 NICU cluster was isolated from an individual in the outpatient department more than 3 months after the first NICU outbreak subsided (Fig 2A).

The genome sequencing data showed that ST1559 isolates were highly clonal and exhibited a lesser degree of phylogenetic diversity than the ST15 isolates, harbouring only 42 SNPs across the core genome of the 19 sequenced isolates. The cases attributable to ST1559 arose later than Outbreak Lineage 1 (between September and December 2012); therefore, we named this group Outbreak Lineage 2.

Assessing the temporal ranges of the outbreak lineages

The 24 Outbreak Lineage 1 strains mainly originated from a range of patient samples in the NICU complex in May and June 2012, which is consistent with the timing of the first peak in *K. pneumoniae* isolations in the microbiology department. To infer the temporal scope of the Outbreak Lineage 1 strains, we estimated the time of the most recent common ancestor (TMRCA) of sequenced Outbreak Lineage 1 isolates using Bayesian phylogenetic methods implemented in BEAST (Fig 2A) (Drummond & Rambaut, 2007). We estimated a median substitution rate of 2.7×10^{-6} substitutions per site per year for Outbreak Lineage 1 (95% highest posterior density (HPD): 6.8×10^{-7}–5.0×10^{-6}), corresponding to an inferred MRCA dating back to early 2011 (95% highest posterior density interval [HPD] 2009–2012) (Fig 2B). However, we interpreted these data with caution noting a relatively weak correlation ($R^2 = 0.134$) between root-to-tip branch lengths and the time of isolation, which was consistent with the outbreak occurring over a short time frame.

To address the limitations of the Bayesian phylogenetic analysis, we screened an additional 412-stored *K. pneumoniae* isolated in the hospital microbiology laboratory between 2005 and 2013 by PCR amplification. To achieve this, we used comparative genomics on a WGS scale to identify unique regions found in the 25 Outbreak

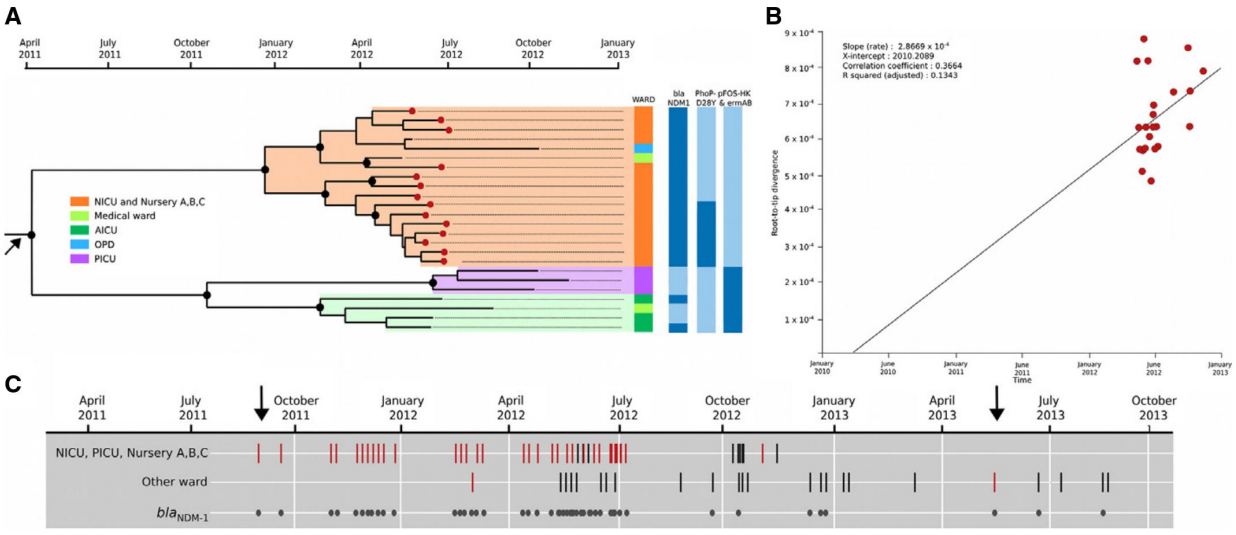


Figure 2. Temporal phylogenetic reconstruction and retrospective PCR detection of *Klebsiella pneumoniae* Outbreak Lineage 1.

A Bayesian phylogenetic reconstruction of Outbreak Lineage 1 using the core genome SNPs of 24 isolates with a corresponding temporal timescale. The black filled circles indicate a posterior probability > 0.8 on internal nodes. The red filled circles indicate *K. pneumoniae* isolates cultured from blood. The ward from which each specimen was collected: orange (NICU and nurseries A, B, C), dark green (AICU), light green (medical ward), light blue (OPD, outpatient department) and purple (PICU). The presence (dark blue) and absence (light blue) of specific genetic determinants for each of the 24 isolates, including bla_{NDM-1}, are highlighted.

B Linear regression plot of the date (*x*-axis) against the root-to-tip divergence (*y*-axis) used to indicate correspondence between phylogenetic and temporal signal. Estimation of temporal signal was performed in Path-O-Gen v1.4 for the core genome SNP-based maximum likelihood phylogeny of 24 Outbreak Lineage 1 *K. pneumoniae*. Additional estimated parameters are inset.

C The temporal distribution of 70 PCR-confirmed Outbreak Lineage 1 *Klebsiella pneumoniae* infections by their ward of isolation. First row: outbreak isolates originating from the NICU, PICU and the nurseries. Second row: isolates originating from other wards, including the AICU, gynaecology ward, maternity ward, medical ward, OPD and surgical ward. The red bars correspond to isolates originating from blood; the black bars correspond to isolates originating from all other specimen types. Third row: the distribution of PCR-confirmed Outbreak Lineage 1 strains containing bla_{NDM-1}. Arrows signify the first and last occasions on which bla_{NDM-1} positive blood-borne *Klebsiella pneumoniae* Outbreak Lineage 1 strains were isolated.

© 2015 The Authors

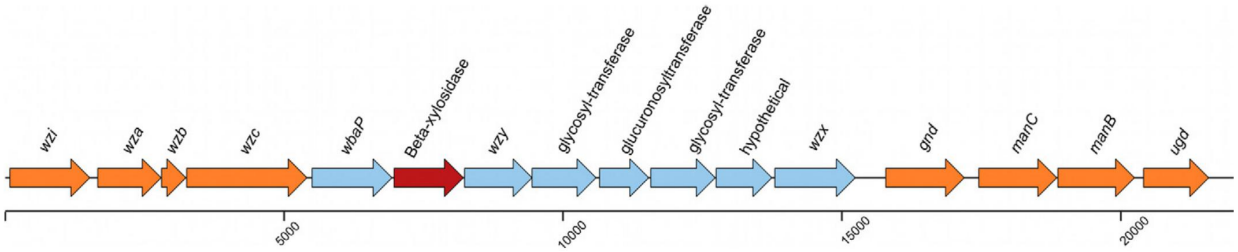


Figure 3. The architecture of the novel capsule biosynthesis cluster (*cps*) in the ST15 Outbreak Lineage 1 *Klebsiella pneumoniae*.
Graphical representation showing the composition of the *cps* gene cluster within the Outbreak Lineage 1 *K. pneumoniae* isolates. The genes are identified by either their common name (e.g. *wba*P) or their predicted function (e.g. beta-xylosidase). Genes that are shared by other strains of *K. pneumoniae* are coloured orange, and genes specific to this cps cluster are coloured blue and red. The beta-xylosidase gene (red) was selected as the target for lineage-specific PCR amplification.

Lineage 1 strains, which included a chromosomal cluster specific for its novel capsule type (Fig 3), as well as a marker for a lineage-specific (pNDM-MAR like) plasmid (see below for details of these regions) (Supplementary Table S2). We additionally included primers for the $bla_{\text{NDM-1}}$ cassette, which was present in 19 of the 25 Outbreak Lineage 1 strains and absent by an apparent perfect excision from pNDM-MAR in the remaining six isolates. By PCR amplification using the specific novel capsular type and pNDM-MAR primers, we identified a further 46 suspect Outbreak Lineage 1 isolates, of which 35 additionally harboured $bla_{\text{NDM-1}}$. These supplemental isolates were largely confined to the NICU, PICU and the nearby nurseries, with the first confirmed isolation dating back to September 2011 (Fig 2C). Despite the limitations of the Bayesian phylogenetic reconstruction, the complementary PCR screening data suggest that Outbreak Lineage 1 was circulating within these wards for at least 6 months prior to recognition of the outbreak and predicted that the lineage likely arose within the hospital in early 2011.

Again, using the WGS data, we identified a similarly informative region within the genome of the Outbreak Lineage 2 isolates to study the pattern of spread of this lineage across the hospital. Resulting PCR amplifications of the Outbreak Lineage 2-specific capsule gene (K14) [see below for details of this region (Table 1)] confirmed that there were likely at least 26 infections caused by this strain (14 cases of bacteraemia) on the NICU, PICU, nurseries and medical wards. A strain belonging to this Outbreak Lineage was first isolated in a medical ward in August 2012, and another was isolated in December after the suspected outbreak in November.

Comparative genomics and lineage-specific gene content

Aiming to understand the genetic basis for the disease phenotypes associated with Outbreak Lineage 1 and Lineage 2, we performed comparative genome analyses by aligning the entire gene contents of the two outbreak lineage strains against the background (non-outbreak) hospital strains. Genetic features that were unique to either outbreak lineage and/or exhibiting potential virulence or antimicrobial resistance were considered as relevant disease markers.

We were able to differentiate the isolates belonging to Outbreak Lineage 1 from all others circulating in this hospital because they possessed a region encoding a previously undescribed capsular antigen biosynthesis (*cps*) cluster which constituted an uncharacterised *K. pneumoniae* capsular K-type (Fig 3). This lineage also possessed a wide complement of additional elements related to virulence and persistence in *Klebsiella* (Podschun & Ullmann, 1998; Schroll *et al*, 2010) (Table 1). While type 1 fimbrial cluster was found in all *K. pneumoniae* strains, this lineage harboured several additional fimbrial and adhesin gene clusters, including a distinctive locus II type six secretion system (T6SS) similar to that found in *K. pneumoniae* 342 (> 99% identity) (Sarris *et al*, 2011) and an intact copy of a genomic island (KR116 KpGI-5) containing a fim2 biofilm associated gene cluster (Van Aartsen *et al*, 2012) (Table 1). Iron acquisition has been strongly implicated in virulence in *K. pneumoniae* (Luck *et al*, 2001; Raymond *et al*, 2003; Ma *et al*, 2005; Lawlor *et al*, 2007). Besides the conventional enterobactin that was found in all strains, Outbreak Lineage 1 again carried a surplus of iron-acquiring systems, including *kfu*, *fec* and yersiniabactin. The yersiniabactin encoding region was identical across Outbreak Lineage 1, distinct from that of the Outbreak Lineage 2 isolates, and appeared to have been co-transferred with additional features as described in other members of the Integrative Conjugative Element (ICE) family (Paauw *et al*, 2010).

We additionally identified several loci associated with persistence and pathogenesis in the Outbreak Lineage 2 strains (Table 1). Of note, we identified multiple loci associated with attachment, including a biofilm encoding *mrk*ABCD type 3 fimbriae locus (Schroll *et al*, 2010) (Table 1). These Outbreak Lineage 2 strains were defined by gene alignment to be capsule type K14. The Outbreak Lineage 2 isolates carried a plasmid-encoded *fec* locus, enterobactin and a yersiniabactin operon. We additionally identified a thermal resistance (temperature survival) gene cluster exhibiting extensive DNA homology to an equivalent region described in *Cronobacter sakazakii* (Gajdosova *et al*, 2011).

Antimicrobial resistance gene content of the Outbreak Lineages

Apart from the virulence-associated genes, we additionally hypothesised that the apparent success of *K. pneumoniae* Outbreak Lineage 1 was further driven by selection induced by antimicrobial usage within the hospital. All Outbreak Lineage 1 isolates were resistant to a broad array of antimicrobials (Fig 1C, Supplementary Table S1). Using the WGS data, we predicted all of the known resistance genes (resistome) found within the sequences (Table 1 and Fig 4). Consistently, the resistome of Outbreak Lineage 1 was extensive and distinct from the other *K. pneumoniae* strains, predicting generic susceptibility to colistin and tigecycline only. All of the Outbreak

© 2015 The Authors

Table 1. The antimicrobial resistance and hypothetical virulence determinants in *Klebsiella pneumoniae* Outbreak Lineages 1 and 2.

Organism	Location	Genetic element	Function	Presence in *K. pneumoniae* strains
ST15 Outbreak Lineage I	pNDM-MAR like	bla_{NDM-1}/*qnrB1*/$bla_{CTX-M-15}$/*catB3/OXA-30/acc(6')-Ib-cr*	carbapenem/fluoroquinolone/cephalosporin, aztreonam/chloramphenicol/oxacillin/aminoglycoside resistance	Outbreak 1 only
	Class I integron of pCTX-M3	*arm*A/*mrs*E-*mph*E/*sul*1/*dhfr*/*aad*A2	Aminoglycoside/macrolide/sulphonamides/trimethoprim/spectinomycin resistance	Outbreak 1 only
	Tn5393c of *Aeromonas* pRAS2	*str*A-*str*B	Streptomycin resistance	Outbreak 1 only
	pKN-LS6 plasmid family	*cus/sil/ars* locus	Copper/silver/arsenite ions resistance	Outbreak 1 only
	Chromosome	Novel capsular biosynthesis cluster (K-type)	Immune escape	Outbreak 1 & 3 other ST15
	Chromosome	Filamentous hemagglutination adhesin locus (ShlB/FhaC/HecB)	Hypothetical adhesins	Outbreak 1 & 5 other ST15
	Chromosome	*K. pneumoniae* KTCC 2242 phage mediated fimbrial cluster	Fimbrial biosynthesis	All ST15
	Chromosome	*K. pneumoniae* 342 T6SS	Host–pathogen interaction and interbacterial communication	All ST15 & all ST14
	Chromosome	ICE-KpNp1	Metal ion uptake systems, yersiniabactin and mobilisation proteins	Outbreak 1 only
	Chromosome	KR116 KpGI-5 genomic island	Biofilm formation	All ST15
	Chromosome	*gyr*A mutation S83F, D87A	Fluoroquinolone resistance	All ST15 + 1 background lineage
	Chromosome	*par*C mutation S80I	Fluoroquinolone resistance	All ST15 + 11 background lineages
ST1559 Outbreak Lineage 2	pKDO1 plasmid family	*sul*2/*str*A-*str*B/TEM-1/CTX-M-15	Sulphanomides/streptomycin/ampicillin/cephalosporins, aztreonam resistance	Found in multiple lineages
	Undetermined plasmid	OXA-30/aac(6')-Ib-cr	Oxacillin/aminoglycoside resistance	Found in multiple lineages
	Undetermined plasmid	*qnr*B1	Fluoroquinolone resistance	Found in multiple lineages
	IS5 mediated, undetermined plasmid	*C. sakazakii* thermoresistance cluster	Hypothetical heat shock resistance	Outbreak 2 only
	pKPN_CZ backbone	*cus/sil* locus	Copper/silver resistance	Outbreak 2 & 4 background lineages
	pCC416	*ars* locus	Arsenite resistance	Outbreak 2 & 3 background lineages
	Undetermined	Extra *ast*ABCDE	Amino acid transport and metabolism	Outbreak 2 only
	Undetermined	IS903 flanked transposome	Amino acid transport and metabolism	Outbreak 2 only
	Chromosome	Yersiniabactin and T4SS	Iron uptake and mobilisation protein	Outbreak 2 only
	Chromosome	*mrk*ABCD locus	Biofilm formation	Outbreak 2 & 4 background lineages

© 2015 The Authors

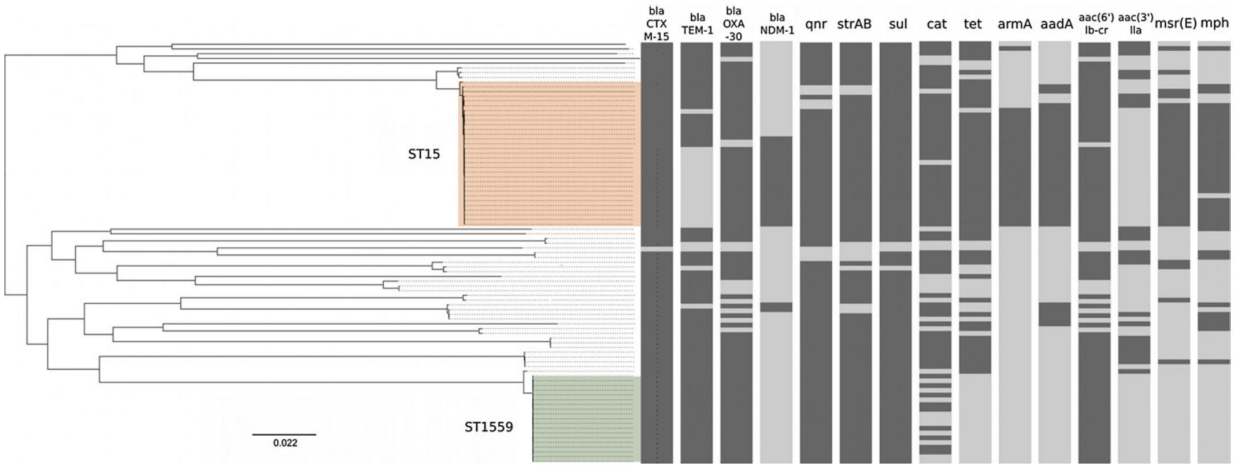


Figure 4. The resistome of the 89 hospital isolates of *Klebsiella pneumoniae*.
Maximum likelihood phylogeny based on core genome SNPs of the 89 *Klebsiella pneumoniae* isolates adapted from Fig 1 set against the resistome of 15 antimicrobial resistance genes (named at the top of the figure). The dark and light grey boxes show the presence and absence of the specific resistance genes in the corresponding isolates, respectively. *bla*, beta-lactamase; *qnr*, *aac(6')-Ib-cr*, quinolone resistance genes; *strAB*, streptomycin resistance genes; *sul*, sulphonamide resistance gene; *cat*, chloramphenicol resistance gene; *tet*, tetracycline resistance gene; *armA*, *aadA*, *aac(3)-IIa*, aminoglycoside resistance genes; and *msr(E)*, *mph*, macrolide resistance genes.

Lineage 1 organisms harboured an MDR plasmid, which shared a high degree of DNA homology (> 99% identity) to the previously described plasmid pNDM-MAR (Villa *et al*, 2012). Within Outbreak Lineage 1, the presence of $bla_{\text{NDM-1}}$ correlated precisely with resistance to carbapenems in isolates originating from the NICU (n = 13), the AICU (n = 2), medical ward (n = 1) and the outpatient department (n = 1). Furthermore, the Outbreak Lineage 1 isolates were found to harbour a pCTX-M3 class I integron encoding resistance to an assortment of antimicrobials including aminoglycosides (*armA*) and macrolides (*mrsE* and *mphE*) (Table 1 and Fig 3). Correspondingly, all of the Outbreak Lineage 1 strains also possessed mutations in *gyr*A (S83F, D87A) and *par*C (S80I), which together with *qnrB1* and *aac(6')-Ib-cr* reduce susceptibility to fluoroquinolones (Hopkins *et al*, 2005). We additionally noted some further ward-specific diversification of Outbreak Lineage 1, with some of the NICU strains having mutations in the *phoP*, of which the impact on polymyxin susceptibility is not yet well understood (Fig 2A). Further, the AICU and PICU Outbreak Lineage I strains gained a pFOS-HK plasmid containing the macrolide resistance genes *erm*A/B (Fig 2A). The Outbreak Lineage 2 strains exhibited less antimicrobial resistance than the Outbreak Lineage 1 strains and were generically more similar to the background hospital isolates; this was reflected in the antimicrobial resistance gene content. The Outbreak Lineage 2 strains harboured three beta-lactamases ($bla_{\text{CTXM-15}}$, $bla_{\text{TEM-1}}$ and bla_{OXA30}), the fluoroquinolone resistance genes *qnrB1* and *aac(6')-Ib-cr* and variable presence of *cat* (chloramphenicol acetyltransferase).

Discussion

Here, we have presented an investigation of two outbreaks of MDR *K. pneumoniae* infections in adjoining high-dependency paediatric units in a hospital in Nepal. The outbreaks were characterised by a high proportion of bloodstream infections and a high case-fatality rate. We used WGS in an attempt to characterise strain circulation and diversity in this hospital, and comparative genomics to identify genetic traits that may be associated with the disease phenotype and the persistence of these strains. The notable advances highlighted in this study, in comparison to previous investigations of hospital outbreaks of *K. pneumoniae,* are the use of the large-scale WGS data to deduce the bacterial isolates involved in the outbreaks and to use these data to retrospectively define the scope of outbreak strains across the hospital over a protracted period. These approaches have permitted us to interrogate these serious hospital outbreaks of *K. pneumoniae* with a degree of resolution that would be inaccessible through conventional methods. The WGS approach has previously been implemented to study hospital outbreaks of organisms such as KPC-producing *K. pneumoniae* in the U.S. (Snitkin *et al*, 2012) and *Staphylococcus aureus* in the U.K. (Harris *et al*, 2013). We think that this study contributes major insights into the epidemiology of *K. pneumoniae*: a major global, and often overlooked, Gram-negative pathogen. Further, this investigation was performed in a resource-limited hospital setting, an example of an environment where hospital outbreaks of *K. pneumoniae* can be associated with high mortality.

Klebsiella pneumoniae is a common community and hospital-acquired pathogen, yet is especially prevalent and vicious in high-risk patients in low-income countries (Escobar Pérez *et al*, 2013; Espedido *et al*, 2013; Poirel *et al*, 2014). As such, *K. pneumoniae* are commonly associated with poor outcomes through respiratory tract infections and bloodstream infections in these locations. This is partly because *K. pneumoniae* are particularly successful in adapting to specific niches through the acquisition of genetic material via horizontal gene transfer (HGT) (Ma *et al*, 2005; Lawlor *et al*, 2007; Fouts *et al*, 2008) and are an extensive reservoir of antimicrobial resistance genes within healthcare settings (Tada *et al*, 2013; Wang *et al*, 2013). The organisms studied here were no exception, with 100% and 65% of the 2012 collection of 270 *K. pneumoniae* exhibiting

© 2015 The Authors

resistance to cefotaxime and ofloxacin, respectively. Further, a resistome analysis of the 89 sequenced *K. pneumoniae* revealed an exhaustive repertoire of antimicrobial resistance genes, thus illustrating the threat that circulating antimicrobial resistant *K. pneumoniae* pose in this hospital (Basnyat, 2014). Indeed, the acquisition of a carbapenemase in an already MDR background appears to have directly contributed to one of the two identified outbreaks.

Our analysis of the WGS data permitted us to identify two discrete *K. pneumoniae* lineages causing two temporally distinct outbreaks of bacteraemia in the high-dependency paediatric wards, named Outbreak Lineage 1 and 2, respectively. The first *K. pneumoniae* outbreak was the largest and the most devastating, causing a minimum of 34 bloodstream infections, of which 28 were fatal. *K. pneumoniae* Outbreak Lineage 1 was an ST15 strain, in which the majority of isolates harboured the carbapenemase gene $bla_{\text{NDM-1}}$, which appears to have become well adapted for sustained transmission within this hospital environment, as evidenced by its prolonged detection prior to the outbreak in 2011 through to the latter part of 2013. The distribution of Outbreak Lineage 1 in the hospital, along with its ward-structured phylogeny, supports a scenario where, following initial introduction into adjacent wards, these strains have undergone a series of independent and highly localised micro-clonal expansions. We speculate that the enhanced virulence potential observed in this outbreak clade was induced by the cumulative introductions of several mobile genetic elements (MGEs). The discrepancy in the scale of the accessory genome between the Outbreak Lineage 1 strains and the non-Outbreak Lineage 1 strains implies that the progenitors of this Outbreak Lineage had access to a wider gene pool than that of other hospital *K. pneumoniae* strains, or alternatively that this strain is particularly receptive to HGT. Indeed, the fixation of the novel capsular antigen predates the diversification of Outbreak Lineage 1 (as evidenced by three novel K-type strains without pNDM-MAR). This novel capsule type may have expedited the initial phase of strain persistence in the hospital through an enhanced capacity for attachment and/or colonisation. The acquisition of the yersiniabactin operon and the pNDM-MAR plasmid were unquestionably pivotal events in the emergence of this strain within the hospital. The active gene acquisition and diversification of some Outbreak Lineage 1 strains during the period of investigation, as exemplified by gain of a predicted novel plasmid and the macrolide resistance (*erm*B) cassette in the AICU and PICU clusters, further supports our hypothesis that this lineage was particularly receptive to genetic material encoding additional phenotypes.

The second *K. pneumoniae* outbreak occurred later in 2012 and, whilst still major, was less severe (with respect to scale) than the first outbreak, causing 26 infections (14 bacteraemia cases with eight fatalities). The Outbreak Lineage 2 strains were also MDR, yet all strains were susceptible to carbapenems and aminoglycosides. However, like Outbreak Lineage 1, Outbreak Lineage 2 also had an expanded accessory genome containing several iron-acquiring systems, including a yersiniabactin distinct to those found in Outbreak Lineage 1. Crucially, siderophores such as yersiniabactin serve the key function of iron acquisition for *K. pneumoniae* in low iron conditions, and yersiniabactin is crucial for replication in the host (Lawlor *et al*, 2007; Bachman *et al*, 2011). We speculate that the excess of these iron acquisition systems contributed greatly to the ability of both outbreak lineages of *K. pneumoniae* to cause systemic infections, potentially by allowing more efficient replication whilst in the bloodstream through enhanced iron chelation.

The endemicity of NDM-1 in South Asia is being facilitated by efficient HGT between differing bacterial species, as exquisitely demonstrated by the presence/absence of $bla_{\text{NDM-1}}$ in Outbreak Lineage 1 and on differing plasmids of other STs in this hospital (Fig 1C). The data suggest that $bla_{\text{NDM-1}}$ is common in healthcare facilities in Nepal. Indeed, whilst this work was in progress, NDM-1-producing *K. pneumoniae* was identified in another major hospital in the Kathmandu Valley (Tada *et al*, 2013). It appears that the strain identified in this study shares some of the traits of Outbreak Lineage 1. Both the strains from the cited study and Outbreak Lineage 1 were of CC14 (specific ST not stated) and both possessed a pCTX-M3 *arm*A integron cassette, $bla_{\text{NDM-1}}$, and had identical fluoroquinolone resistance inducing mutations (S83F, D87A in *gyr*A and S80I in *par*C). Whilst the genetic environment of $bla_{\text{NDM-1}}$ was not specified in the previous study, it is possible that these two clades/clones are closely related. We suspect that these strains have covertly spread throughout a number of healthcare facilities in the region; detailed and broader follow-up work and phylogenetic reconstruction are clearly warranted to understand the true extent of the temporal and geographical circulation of these strains in Nepal.

Our work contains some limitations that should be considered in the context of this investigation. Firstly, the strains subjected to WGS were collected over a relatively short time frame, with our analysis concentrated on strains falling within the two outbreak peaks. This approach restricted our investigative power over a more extensive timescale. Bayesian phylogenetic inference suggests that Outbreak Lineage 1 likely emerged in 2011. This date was consistent with our clade-specific PCR amplification results; however, the timescale of sampling and rapidity of these outbreaks offer only limited variation across the core genome with which to infer evolutionary and epidemiological processes, and thus may impact the robustness of this analysis. Further, we did not exhaustively perform WGS on all *K. pneumoniae* isolated in the hospital in 2012. However, we argue that our strain selection and supplementary PCR analysis were sufficient to estimate the extent of the outbreaks. Of course, we cannot guarantee that all outbreak strains inferred by PCR amplification were closely related, as detection of the selected loci may be compromised by HGT within the circulating Gram-negative population. However, these capsular and antimicrobial resistance genes were relevant in this location, and such targets have been previously exploited for a similar purpose (Villa *et al*, 2012; Chen *et al*, 2014). Since the time frame for this study was short, we predict that the impact of capsule misidentification due to HGT in the two highlighted lineages was minimal.

Using an approach similar to the one described here, the genomic diversity and spread of KPC-producing ST258 *K. pneumoniae* in hospitals in the U.S. has been recently investigated (Deleo *et al*, 2014; Wright *et al*, 2014). Further, a *cps* gene cluster introduction into these ST258 strains has been shown to facilitate substantial divergence within this ST, and retrospective PCR based on distinct capsular markers has been utilised to probe its expansion across U.S. healthcare facilities (Chen *et al*, 2014). While phylogenetics and comparative genomics have been performed previously for outbreaks of *K. pneumoniae*, such studies have not typically considered the endemic background strain diversity and have focussed on specific problematic lineages in developed countries (Snitkin *et al*, 2012; Ramos *et al*, 2014). Our work is novel in its approach to

 © 2015 The Authors

utilising genomic data to explore highly fatal nosocomial outbreaks of *K. pneumoniae* in a low-income setting. Most significantly, we identified the emergence of a capsule switching NDM-1-bearing ST15 strain in a single hospital. NDM-1-producing ST15 *K. pneumoniae* has been reported repeatedly (Poirel *et al*, 2011; Arpin *et al*, 2012; Hu *et al*, 2013), but its transmission ability and pathogenic capacity have not been assessed. We propose that further studies should concentrate on the diversity and spread of this specific clone, particularly in healthcare settings in developing countries.

In conclusion, we have spatiotemporally characterised invasive *K. pneumoniae* populations in one hospital in Nepal using WGS. These data permitted a comprehensive understanding of the emergence of two highly virulent clones, which could then be retrospectively identified using relevant lineage-specific genetic determinants. Owing to the severity and the mortality of these outbreaks, it is regrettable that some form of routine genetic characterisation was not performed in real time. After the first cases of *K. pneumoniae* bacteraemia, NDM-1 PCR amplification and MLST would have added some vital insights into these outbreaks in a clinically relevant time frame and may have reduced their magnitude. We suggest that the early detection of a specific NDM-1 clone on the NICU complex in 2011 would have alerted hospital staff to intervene, permitted on-going surveillance with a review of the antimicrobial treatment policy and sustained screening and patient isolation. The added benefit of WGS was to dissect these outbreaks in extremely fine detail. We argue that some form of real-time genetic characterisation, and clade-specific PCR amplification, should undoubtedly factor into future routine microbiology in such hospitals to rapidly provide data required to intervene in such outbreaks in a timely manner.

Materials and Methods

Ethics statement

The institutional review boards of the Patan Academy of Health Sciences (PHAS) and the Nepal Health Research Council (NHRC) provided ethical approval for this study.

Setting

Patan Hospital is a 450-bed government hospital providing emergency and elective outpatient and inpatient services located within the Kathmandu Valley in the Sub-Metropolitan City of Lalitpur, Nepal. This study describes a *K. pneumoniae* outbreak in this hospital in 2012. The outbreak appeared to be restricted to the three nurseries (A, B and C), the NICU and the PICU on the first floor and the medical ward and the adult intensive care unit (AICU) directly above on the second floor (as described in Fig 1A). The NICU is a seven-bed ward (one bed for isolation) for neonates (≤ 28 days old) requiring intensive therapy. The ward typically treats babies with intrauterine growth restriction (IUGR) or premature neonates with sepsis or requiring ventilation. It has seventeen full-time medical staff (eight nurses and nine doctors) with four staff on shift at any one time. The PICU is a six-bed ward for children > 28 days old requiring intensive therapy. The ward typically treats children with shock, most often due to sepsis and those requiring mechanical ventilation. The PICU has nineteen full-time medical staff (ten nurses and nine doctors). The nurseries are adjacent to the maternity ward and the NICU and PICU, and newborn children are routinely transferred into them for general observations. Combined, these nurseries contain twenty beds. The doctors regularly move between the NICU, PICU and nurseries, but the nurses are restricted to their respective wards. The second floor AICU is an eight-bed ward for adults (> 18 years) requiring intensive therapy, typically those presenting with shock, cardiac emergencies and those requiring mechanical ventilation. The AICU has eight full-time medical staff (four nurses and four doctors) and is located adjacent to a general medical ward containing sixty beds (two for isolation) for patients with a variety of conditions. Data relating to individual patient treatment regimes and clinical variables (other than death) were not accessible.

Bacteriology and antimicrobial susceptibility testing

The isolates described in this study were obtained from cultures of clinical specimens taken as part of routine clinical care from Patan Hospital inpatients between May and December 2012. Microorganisms were isolated and identified using routine microbiological methods. Briefly, blood cultures were processed using the BD BACTEC™ (Becton Dickinson, UK) Instrumented Culture System incubating BACTEC™ Peds Plus™/F culture vials containing 1–2 ml of peripheral blood for children. Adult blood samples (typically > 5 ml) were collected in BACTEC™ Plus Aerobic/F Culture Vials. Samples flagged as positive by the BD BACTEC™ system were subcultured onto MacConkey and blood agar plates and incubated at 37°C to isolate the infecting organism. All other clinical specimens (e.g. tracheal aspirates, wound swabs, urine) were processed according to standard microbiological procedures. *K. pneumoniae* (and other Gram-negative organisms) was identified using standard biochemical tests and API20E testing, following the manufacturer's instructions (Bio Merieux, France).

Antimicrobial susceptibilities were tested at the time of isolation by the modified Bauer-Kirby disc diffusion method, as recommended by the CLSI guidelines (CLSI, 2012). Mueller-Hinton agar and antimicrobial discs were purchased from Unipath, Basingstoke, United Kingdom. *Escherichia coli* ATCC 25922 was used as the control strain. The inhibitory zone sizes were recorded and interpreted according to current CLSI breakpoint guidelines (CLSI, 2012). The following antimicrobials were used for *K. pneumoniae* susceptibility testing: amoxicillin (AMX), cefotaxime (CTX), ciprofloxacin (CIP), ofloxacin (OFX), trimethoprim/sulfamethoxazole (SXT), gentamicin (GEN), amikacin (AMK), chloramphenicol (CHL), meropenem (MEM) and imipenem (IPM).

Whole-genome sequencing

We selected a random subset of 90 *K. pneumoniae* (using a random integer generator) from a collection of 135 *K. pneumoniae* that were isolated and stored in Patan Hospital between May and December 2012. This collection contained *K. pneumoniae* isolated from blood (n = 28), urine (n = 27), catheter and CV tips (n = 32), respiratory tract infections (n = 20), pus (n = 14) and other infection sites (n = 14) (selected isolated outlined in Supplementary Table S1). These organisms were subjected to DNA extraction by using the Wizard Genomic DNA Extraction Kit (Promega, Wisconsin, USA), and 2 μg was subjected to whole-genome sequencing (WGS) on an

© 2015 The Authors

Illumina HiSeq2000 Platform (Illumina, San Diego, USA) to generate 150 bp paired-end reads, as previously described (Quail *et al*, 2008). Reads were mapped to the reference *K. pneumoniae* MGH78578 by using BWA, and QC stats were exported to graphs and manually assessed (Li & Durbin, 2010). The average Phred quality score of the all sequencing read sets was 35 (33.7–37). *De novo* assemblies were generated for each read set using Velvet (version 1.2.03) and VelvetOptimiser, with each read set mapped back to each assembly (Zerbino & Birney, 2008). The resulting *K. pneumoniae* short-read sequence data were deposited in the European Read Archive under the Accession Numbers ERR349747 to ERR349855 inclusive. The assemblies of representatives of two outbreak lineages are available at GENBANK under Accession numbers CDQG01000001-CDQG01000097 (Outbreak Lineage 1) and CDPE01000001-CDPE01000062 (Outbreak Lineage 2), and the novel capsule biosynthesis gene from Outbreak Lineage 1 is available under the Accession Number LN714331.

Genome annotation, core genome construction and phylogenetic analysis

For each *de novo* assembly, coding sequences were predicted and annotated using the rapid prokaryotic genome annotation tool, Prokka (Seemann, 2014). Predicted coding sequences comprising more than 5% undetermined nucleotides or shorter than 120 nucleotides in length were excluded from further analysis. Coding sequences were translated into proteins, and protein clustering was performed iteratively by CD-hit (Li & Godzik, 2006) initially with 100% identity and 100% length match, then reducing by 0.1% until reaching a minimum 98% identity. The "core" *K. pneumoniae* genome was defined as concatenation of coding sequences present in one copy in all of the final 89 (one failure) *K. pneumoniae* assemblies. We used this core genome alignment to construct a maximum likelihood phylogeny of the 89 *K. pneumoniae* isolates using RAxML after removing predicted regions of recombination by using Gubbins (Stamatakis, 2006; Croucher *et al*, 2014). In brief, Gubbins predicts and removes recombinant regions using a spatial scanning statistic to identify genomic regions of significantly raised SNP density on each branch of the tree. Once recombinant regions are removed, the tree is recalculated and the process repeated until the tree or recombination predictions are the same as a previous iteration or when a user-specified maximum number of iterations are reached. Specific loci of interest were retrieved from the annotated assemblies, aligned with MUSCLE (Edgar, 2004) and visualised in SeaView (Gouy *et al*, 2010). Multilocus sequence typing (MLST) was performed for all *K. pneumoniae* isolates by extracting the seven typing genes *in silico* (*gapA*, *infB*, *mdh*, *pgi*, *phoE*, *rpoB*, *tonB*) (Diancourt *et al*, 2005). Capsule genotyping was performed by comparing the sequence of the *wzc* gene to those deposited in Genbank (Pan *et al*, 2013). The resistance gene profiles and content (resistome) of each isolate were characterised using a manually curated database, based on the ResFinder database (Zankari *et al*, 2012). Each gene in the database was mapped against the isolate assemblies to identify complete genes. Where fragments matching the 5′ or 3′ ends of resistance genes were identified at contig boundaries, sequencing reads were mapped to all matching candidate genes and their presence assessed based on mapping coverage across the gene.

Accessory genome assembly

To improve the assembly of accessory regions, sequencing reads for each isolate were mapped against the *K. pneumoniae* 1084 reference genome using SMALT (https://www.sanger.ac.uk/resources/software/smalt/) (Lin *et al*, 2012). Unmapped read pairs were extracted and assembled as described above. Contigs were discarded if they were shorter than 1,000 bp or contained a region of at least 1,000 bp that matched the reference genome using BLAST. The remaining contigs from all isolates were compared in a pairwise manner using BLAST. In each comparison, contigs were retained if they contained at least 1,000 bp of unmatched sequence, or if they were the longer of the two compared sequences. When two or more contigs were identical, only one was retained and all other contigs were discarded. The presence or absence of contigs in the remaining accessory genome was assessed for each isolate by mapping sequencing reads against the reference genome and accessory contigs simultaneously.

Comparative genomics

The pan-genome of 89 *K. pneumoniae* strains constructed by the accessory genome assembly method described above was used for comparative genomic analysis to identify genetic features distinct to the two outbreak lineages (present in either outbreak lineage but absent in almost all remaining strains). These identified genetic features were individually assessed using BLAST to determine their origin, location within the genome (plasmid or chromosomal borne) and function based on homology to the highest score BLAST hits. Poorly characterised, putative prophages and functionally unknown regions were excluded, and regions for which annotations matched with known potential virulence factors were retained for further analysis.

Bayesian phylogenetic methods

To investigate the timescale of the emergence and spread of these *K. pneumoniae* outbreaks, we inferred a maximum likelihood phylogeny for 24 *K. pneumoniae* isolates which represented the primary outbreak strain (described herein as Outbreak Lineage 1) using PhyML under the HKY85 + Γ_4 model of substitution (Guindon *et al*, 2010). To test for temporal signal in the phylogeny, we assessed the correlation between root-to-tip distance and the date of isolation (day, month, year) in Path-O-Gen (v1.4). A Bayesian Markov chain Monte Carlo (MCMC) approach was then applied in the BEAST package (v1.8) to infer the posterior distribution of trees and determine dates of divergence of lineages of interest (Drummond & Rambaut, 2007). We used an HKY85 + Γ_4 substitution model, a relaxed lognormal molecular clock, a constant size demographic model and tip dates defined as day, month and year of isolation; Bayes factor testing determined this model to be the best fit of various models tested. Analyses were run in triplicate using a continuous 200 million generation MCMC chain with samples taken every 20,000 generations. All analyses were run until parameters reached convergence with ESS > 200, as assessed using Tracer (v1.5). Triplicate runs were combined using LogCombiner (v1.8) with removal of 20% for burn-in.

 © 2015 The Authors

The paper explained

Problem
Klebsiella pneumoniae is a Gram-negative bacterium that colonises humans and animals. The organism is adept at gaining multiple antimicrobial resistance genes and, given the right opportunities, can cause serious infections in susceptible individuals. In 2012 there was an outbreak of highly drug resistant *Klebsiella pneumoniae* bloodstream infections on high-dependency wards for infants in our hospital in Kathmandu, Nepal. This outbreak had two peaks within the year and a mortality of 75%.

Results
We selected many of these outbreak strains and some of non-outbreak *Klebsiella pneumoniae* from across the hospital and performed whole-genome sequencing. Our analysis revealed two independent outbreak strains of *Klebsiella pneumoniae* that we named Outbreak Lineage 1 and Outbreak Lineage 2. We were able to use these data to reconstruct these outbreaks through time and across the affected wards with meticulous accuracy. We found that strains belonging to Outbreak Lineage 1 were circulating in the hospital for 6 months before the outbreak and had gained genes facilitating infection and antimicrobial resistance.

Impact
This work allows a comprehensive insight into how hospital outbreaks of drug resistant *Klebsiella pneumoniae* occur in low-income settings. The power of these data outlines that some real-time genetic characterization should be performed during future healthcare infection control practices in both high- and low-income settings to identify outbreaks and limit onwards transmission.

Retrospective PCR screening

To retrospectively screen for organisms belong to Outbreak Lineage 1 and Outbreak Lineage 2, we identified *K. pneumoniae* lineage-specific regions using comparative genomics, as described above. Primers were designed (or selected from previous publications) for lineage-specific genes in the capsule regions, the $bla_{\text{NDM-1}}$ cassette and the corresponding MDR plasmid. These primers, outlined in Supplementary Table S2, were combined to form a multiplex with predicted amplification sizes of 400, 600, 1,000 and 1,300 bp. PCR amplification was performed on the 412 *K. pneumoniae* that were isolated and stored in Patan Hospital between 2005 and 2013. DNA was extracted by inoculating the cultured *K. pneumoniae* into 50 μl of sterile water and exposing them to 100°C using a microtube heating block (Stuart equipment, UK) for 5 min. Extractions were centrifuged briefly and 2 μl of supernatant was subjected to multiplex PCR amplification for 5 min at 94°C, 35 cycles of 30 s at 94°C, 30 s at 55°C and 105 s at 72°C, followed by 7 min at 72°C. PCR was performed using FastStart Taq DNA polymerase (Roche, Switzerland) and icycler thermal cycler (Bio-Rad, USA). PCR amplifications were UV visualised after agarose gel electrophoresis and staining with ethidium bromide.

For more information

http://www.cdc.gov/drugresistance/threat-report-2013/pdf/ar-threats-2013-508.pdf
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/244058/20130902_UK_5_year_AMR_strategy.pdf
http://www.pasteur.fr/recherche/genopole/PF8/mlst/primers_Kpneumoniae.html
https://www.sanger.ac.uk/resources/software/
http://www.oucru.org/enterics/
http://www.sciencedaily.com/releases/2008/12/081201105706.htm

Supplementary information for this article is available online:
http://embomolmed.embopress.org

Acknowledgements
The authors wish to thank all the unit staff at the Patan Hospital in Kathmandu for assisting in sample collection, data collection and patient care. The Wellcome Trust of Great Britain supported this work. Stephen Baker is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (100087/Z/12/Z). CJB and AKC are supported by the Medical Research Council (grant number G1100100).

Author contributions
HCT, AK, DPT, CJB, AKC, ME, KB, SD, TJ, TLTP, NTDH and THT performed experiments CJB, AKC, ME, KB, SD, CT, SRH, TJ, MAR and NRT provided methods or reagents for this work. AK, SD, CT, SS, SJ, BB and GT provided clinical input and data for this work. HCT, AK, DPT, CJB, AKC, ME, KB, CT, SRH, TJ and MAR performed analysis for this work. HCT, AK, GT, NRT and SB wrote the first draft of the manuscript. HCT, AK, DPT, CJB, AKC, ME, KB, SD, CT, SRH, TJ, TLTP, NTDH, THT, SS, SJ, BB, GT, NRT, MAR and SB contributed to and read the final version of the manuscript.

Conflict of interest
The authors declare that they have no conflict of interest.

References

Arpin C, Noury P, Boraud D, Coulange L, Manetti A, André C, M'Zali F, Quentin C (2012) NDM-1-producing Klebsiella pneumoniae resistant to colistin in a French community patient without history of foreign travel. *Antimicrob Agents Chemother* 56: 3432 – 3434

Bachman MA, Oyler JE, Burns SH, Caza M, Lépine F, Dozois CM, Weiser JN (2011) Klebsiella pneumoniae yersiniabactin promotes respiratory tract infection through evasion of lipocalin 2. *Infect Immun* 79: 3309 – 3316

Basnyat B (2014) Antibiotic resistance needs global solutions. *Lancet Infect Dis* 14: 549 – 550

Bialek-Davenet S, Criscuolo A, Ailloud F, Passet V, Jones L, Garin B, Hello SL, Arlet G, Decré D, Brisse S (2014) Genomic definition of hypervirulent and multidrug-resistant Klebsiella pneumoniae clonal groups. *Emerg Infect Dis* 20: 1812 – 1820

Brisse S, Fevre C, Passet V, Issenhuth-Jeanjean S, Tournebize R, Diancourt L, Grimont P (2009) Virulent clones of Klebsiella pneumoniae: identification and evolutionary scenario based on genomic and phenotypic characterization. *PLoS ONE* 4: e4982

Centres for Disease Control and Prevention (2013) Antibiotic Resistance Threats in the United States, 2013.

Chen L, Chavda KD, Findlay J, Peirano G, Hopkins K, Pitout JDD, Bonomo RA, Woodford N, DeLeo FR, Kreiswirth BN (2014) Multiplex PCR for identification of two capsular types in epidemic KPC-producing Klebsiella pneumoniae sequence type 258 strains. *Antimicrob Agents Chemother* 58: 4196 – 4199

CLSI (2012) Clinical and Laboratory Standards Institute: Performance standards for antimicrobial susceptibility testing; twentieth informational supplement.

© 2015 The Authors

Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR (2014) Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 44: 1 – 13

Deleo FR, Chen L, Porcella SF, Martens CA, Kobayashi SD, Porter AR, Chavda KD, Jacobs MR, Mathema B, Olsen RJ *et al* (2014) Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 Klebsiella pneumoniae. *Proc Natl Acad Sci USA* 111: 4988 – 4993

Diancourt L, Passet V, Verhoef J, Patrick AD, Grimont PAD, Brisse S (2005) Multilocus sequence typing of Klebsiella pneumoniae nosocomial isolates. *J Clin Microbiol* 43: 4178 – 4182

Drummond AJ, Rambaut A (2007) BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7: 214

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792 – 1797

Escobar Pérez JA, Olarte Escobar NM, Castro-Cardozo B, Valderrama Márquez IA, Garzón Aguilar MI, Martinez de la Barrera L, Barrero Barreto ER, Marquez-Ortiz RA, Moncada Guayazán MV, Vanegas Gómez N (2013) Outbreak of NDM-1-producing Klebsiella pneumoniae in a neonatal unit in Colombia. *Antimicrob Agents Chemother* 57: 1957 – 1960

Espedido BA, Steen JA, Ziochos H, Grimmond SM, Cooper MA, Gosbell IB, van Hal SJ, Jensen SO (2013) Whole genome sequence analysis of the first Australian OXA-48-producing outbreak-associated Klebsiella pneumoniae isolates: the resistome and in vivo evolution. *PLoS ONE* 8: e59920

Fouts DE, Tyler HL, DeBoy RT, Daugherty S, Ren Q, Badger JH, Durkin AS, Huot H, Shrivastava S, Kothari S *et al* (2008) Complete genome sequence of the N2-fixing broad host range endophyte Klebsiella pneumoniae 342 and virulence predictions verified in mice. *PLoS Genet* 4: e1000141

Gajdosova J, Benedikovicova K, Kamodyova N, Tothova L, Kaclikova E, Stuchlik S, Turna J, Drahovska H (2011) Analysis of the DNA region mediating increased thermotolerance at 58°C in Cronobacter sp. and other enterobacterial strains. *Antonie Van Leeuwenhoek* 100: 279 – 289

Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27: 221 – 224

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59: 307 – 321

Harris SR, Cartwright EJP, Török ME, Holden MTG, Brown NM, Ogilvy-Stuart AL, Ellington MJ, Quail MA, Bentley SD, Parkhill J *et al* (2013) Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. *Lancet Infect Dis* 13: 130 – 136

Hopkins KL, Davies RH, Threlfall EJ (2005) Mechanisms of quinolone resistance in Escherichia coli and Salmonella: recent developments. *Int J Antimicrob Agents* 25: 358 – 373

Hrabák J, Empel J, Bergerová T, Fajfrlík K, Urbásková P, Kern-Zdanowicz I, Hryniewicz W, Gniadkowski M (2009) International clones of Klebsiella pneumoniae and Escherichia coli with extended-spectrum beta-lactamases in a Czech hospital. *J Clin Microbiol* 47: 3353 – 3357

Hu L, Zhong Q, Tu J, Xu Y, Qin Z, Parsons C, Zhang B, Hu X, Wang L, Yu F *et al* (2013) Emergence of blaNDM-1 among Klebsiella pneumoniae ST15 and novel ST1031 clinical isolates in China. *Diagn Microbiol Infect Dis* 75: 373 – 376

Hudson CM, Bent ZW, Meagher RJ, Williams KP (2014) Resistance Determinants and Mobile Genetic Elements of an NDM-1-Encoding Klebsiella pneumoniae Strain. *PLoS ONE* 9: e99209

Lawlor MS, O'connor C, Miller VL (2007) Yersiniabactin is a virulence factor for Klebsiella pneumoniae during pulmonary infection. *Infect Immun* 75: 1463 – 1472

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658 – 1659

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26: 589 – 595

Lin A-C, Liao T-L, Lin Y-C, Lai Y-C, Lu M-C, Chen Y-T (2012) Complete genome sequence of Klebsiella pneumoniae 1084, a hypermucoviscosity-negative K1 clinical strain. *J Bacteriol* 194: 6316

Luck SN, Turner SA, Rajakumar K, Sakellaris H, Adler B (2001) Ferric dicitrate transport system (Fec) of Shigella flexneri 2a YSH6000 is encoded on a novel pathogenicity island carrying multiple antibiotic resistance genes. *Infect Immun* 69: 6012 – 6021

Ma L-C, Fang C-T, Lee C-Z, Shun C-T, Wang J-T (2005) Genomic heterogeneity in Klebsiella pneumoniae strains is associated with primary pyogenic liver abscess and metastatic infection. *J Infect Dis* 192: 117 – 128

Munoz-Price LS, Poirel L, Bonomo RA, Schwaber MJ, Daikos GL, Cormican M, Cornaglia G, Garau J, Gniadkowski M, Hayden MK *et al* (2013) Clinical epidemiology of the global expansion of Klebsiella pneumoniae carbapenemases. *Lancet Infect Dis* 13: 785 – 796

Nordmann P, Naas T, Poirel L (2011) Global spread of carbapenemase-producing enterobacteriaceae. *Emerg Infect Dis* 17: 1791 – 1798

Paauw A, Leverstein-van Hall MA, Verhoef J, Fluit AC (2010) Evolution in quantum leaps: multiple combinatorial transfers of HPI and other genetic modules in Enterobacteriaceae. *PLoS ONE* 5: e8662

Pan Y-J, Lin T-L, Chen Y-H, Hsu C-R, Hsieh P-F, Wu M-C, Wang J-T (2013) Capsular types of klebsiella pneumoniae revisited by wzc sequencing. *PLoS ONE* 8: e80670

Podschun R, Ullmann U (1998) Klebsiella spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clin Microbiol Rev* 11: 589 – 603

Poirel L, Benouda A, Hays C, Nordmann P (2011) Emergence of NDM-1-producing Klebsiella pneumoniae in Morocco. *J Antimicrob Chemother* 66: 2781 – 2783

Poirel L, Yilmaz M, Istanbullu A, Arslan F, Mert A, Bernabeu S, Nordmann P (2014) Spread of NDM-1-producing Enterobacteriaceae in a neonatal intensive care unit in Istanbul, Turkey. *Antimicrob Agents Chemother* 58: 2929 – 2933

Quail M, Kozarewa I, Smith F, Scally A (2008) A large genome center's improvements to the Illumina sequencing system. *Nat Methods* 5: 1005 – 1010

Ramos PIP, Picão RC, De Almeida LGP, Lima NCB, Girardello R, Vivan ACP, Xavier DE, Barcellos FG, Pelisson M, Vespero EC *et al* (2014) Comparative analysis of the complete genome of KPC-2-producing Klebsiella pneumoniae Kp13 reveals remarkable genome plasticity and a wide repertoire of virulence and resistance mechanisms. *BMC Genom* 15: 54

Raymond KN, Dertz EA, Kim SS (2003) Enterobactin: an archetype for microbial iron transport. *Proc Natl Acad Sci USA* 100: 3584 – 3588

Sarris PF, Zoumadakis C, Panopoulos NJ, Scoulica EV (2011) Distribution of the putative type VI secretion system core genes in Klebsiella spp. *Infect Genet Evol* 11: 157 – 166

Schroll C, Barken KB, Krogfelt KA, Struve C (2010) Role of type 1 and type 3 fimbriae in Klebsiella pneumoniae biofilm formation. *BMC Microbiol* 10: 179

Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30: 2068 – 2069

 © 2015 The Authors

Snitkin ES, Zelazny AM, Thomas PJ, Stock F, Henderson DK, Palmore TN, Segre JA (2012) Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. *Sci Transl Med* 4: 148ra116

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688 – 2690

Tada T, Miyoshi-Akiyama T, Dahal RK, Mishra SK, Ohara H, Shimada K, Kirikae T, Pokhrel BM (2013) Dissemination of multidrug-resistant Klebsiella pneumoniae clinical isolates with various combinations of carbapenemases (NDM-1 and OXA-72) and 16S rRNA methylases (ArmA, RmtC and RmtF) in Nepal. *Int J Antimicrob Agents* 42: 372 – 374

Tzouvelekis LS, Markogiannakis A, Psichogiou M, Tassios PT, Daikos GL (2012) Carbapenemases in Klebsiella pneumoniae and other Enterobacteriaceae: an evolving crisis of global dimensions. *Clin Microbiol Rev* 25: 682 – 707

UK Department of Health & Department for Environment Food & Rural Affairs (2013) UK Five Year Antimicrobial Resistance Strategy 2013 to 2018.

Van Aartsen JJ, Stahlhut SG, Harrison EM, Crosatti M, Ou H-Y, Krogfelt KA, Struve C, Rajakumar K (2012) Characterization of a novel chaperone/usher fimbrial operon present on KpGI-5, a methionine tRNA gene-associated genomic island in Klebsiella pneumoniae. *BMC Microbiol* 12: 59

Villa L, Poirel L, Nordmann P, Carta C, Carattoli A (2012) Complete sequencing of an IncH plasmid carrying the blaNDM-1, blaCTX-M-15 and qnrB1 genes. *J Antimicrob Chemother* 67: 1645 – 1650

Walsh TR, Weeks J, Livermore DM, Toleman MA (2011) Dissemination of NDM-1 positive bacteria in the New Delhi environment and its implications for human health: an environmental point prevalence study. *Lancet Infect Dis* 11: 355 – 362

Wang Q, Li B, Tsang AKL, Yi Y, Woo PCY, Liu CH (2013) Genotypic analysis of Klebsiella pneumoniae isolates in a Beijing Hospital reveals high genetic diversity and clonal population structure of drug-resistant isolates. *PLoS ONE* 8: e57091

Wright MS, Perez F, Brinkac L, Jacobs MR, Kaye K, Cober E, van Duin D, Marshall SH, Hujer AM, Rudin SD *et al* (2014) Population structure of KPC-producing Klebsiella pneumoniae isolates from midwestern U.S. hospitals. *Antimicrob Agents Chemother* 58: 4961 – 4965

Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV (2012) Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 67: 2640 – 2644

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18: 821 – 829

License: This is an open access article under the terms of the Creative Commons Attribution 4.0 License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

The genomic signatures of *Shigella* evolution, adaptation and geographical spread

Hao Chung The[1], Duy Pham Thanh[1], Kathryn E. Holt[2,3], Nicholas R. Thomson[4,5] and Stephen Baker[1,5,6]

Abstract | *Shigella* spp. are some of the key pathogens responsible for the global burden of diarrhoeal disease. These facultative intracellular bacteria belong to the family Enterobacteriaceae, together with other intestinal pathogens, such as *Escherichia coli* and *Salmonella* spp. The genus *Shigella* comprises four different species, each consisting of several serogroups, all of which show phenotypic similarity, including invasive pathogenicity. DNA sequencing suggests that this similarity results from the convergent evolution of different *Shigella* spp. founders. Here, we review the evolutionary relationships between *Shigella* spp. and *E. coli*, and we highlight how the genomic plasticity of these bacteria and their acquisition of a distinctive virulence plasmid have enabled the development of such highly specialized pathogens. Furthermore, we discuss the insights that genotyping and whole-genome sequencing have provided into the phylogenetics and intercontinental spread of *Shigella* spp.

[1]*The Hospital for Tropical Diseases, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, 764 Vo Van Kiet, Quan 5, Ho Chi Minh City, Vietnam.*
[2]*Centre for Systems Genomics, University of Melbourne.*
[3]*Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Melbourne, Victoria 3052, Australia.*
[4]*Bacterial Genomics, The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK.*
[5]*Centre for Tropical Medicine and Global Health, University of Oxford, Old Road Campus, Roosevelt Drive, Oxford OX3 7FZ, UK.*
[6]*Department of Pathogen and Molecular Biology, The London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, UK.*
Correspondence to S.B.
sbaker@oucru.org

doi:10.1038/nrmicro.2016.10
Published online 29 Feb 2016

Diarrhoea remains one of the main causes of mortality in young children in low-income countries[1,2]. Although the number of children aged <5 years who die owing to diarrhoea has declined steadily over the past decade, the incidence of diarrhoeal disease has remained comparatively stable over the same period, at ~2.9 episodes per child per year in 2010 (REFS 1,3). In 2012, the WHO estimated that diarrhoeal disease contributes ~3.6% of the global burden of disease in disability adjusted life years (DALY) and results in ~1.5 million deaths annually.

The recent Global Enteric Multicenter Study (GEMS), an expansive case–control study of moderate-to-severe paediatric diarrhoeal disease, identified enterotoxigenic *Escherichia coli* (ETEC) and *Shigella* spp. as the most common bacterial pathogens in sub-Saharan Africa and South Asia[4,5], and found *Shigella* spp. to be the most prevalent pathogens among children 24–59 months old[5]. Historical data suggest that there were ~165 million cases of shigellosis annually mainly in low-income countries and in children <5 years old between 1966 and 1997, resulting in 1.1 million deaths worldwide[6]. More recently, it was estimated that *Shigella* spp. cause ~125 million disease cases annually[7], and that the incidence of shigellosis is 13.2 cases per 1,000 children per year in children aged <5 years in Asia[8]. Importantly, despite the continued high incidence of *Shigella* spp. infections, the mortality rate per case has dropped by 98%; this is probably due to the disappearance of epidemics associated with the highly pathogenic species *Shigella dysenteriae* and improved, more rapid treatment[7].

The Gram-negative bacterial genus *Shigella* belongs to the family Enterobacteriaceae, which also encompasses other enteric pathogens, including ETEC, enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC; also known as Shiga toxin-producing *E. coli* (STEC)), enteroaggregative *E. coli* (EAEC), diffusely adherent *E. coli* (DAEC) and enteroinvasive *E. coli* (EIEC)[9,10]. The pathogenesis and epidemiology of each of these *E. coli* pathovars are distinct and complex, and reflect the diverse catalogue of phenotypic traits that *E. coli* has acquired during its evolution from commensal to pathogen in humans and other mammals[11]. However, *Shigella* spp. stand out from other Enterobacteriaceae: their evolutionary history, mechanism of pathogenesis and human-restricted nature make them unique.

Shigella spp. are intracellular pathogens and are transmitted through the faecal–oral route. They can induce a symptomatic infection via an exceptionally low infectious dose (<10 bacteria), as opposed to *Salmonella* spp. and the various diarrhoeagenic *E. coli* pathovars, which have infectious doses of at least four orders of magnitude greater[12]. *Shigella* spp. cause bacillary dysentery, a severe form of diarrhoea in which blood and mucus can be observed in the stool as a consequence of epithelial cell

© 2016 Macmillan Publishers Limited. All rights reserved.

damage in the lower gut (BOX 1). The highly pathogenic and exotoxin-producing species *S. dysenteriae* was first described in 1897 by Kiyoshi Shiga, who isolated this species from the stool sample of a patient with epidemic dysentery in Japan[13,14]. The genus was expanded soon after: *Shigella flexneri* was identified in 1899, *Shigella sonnei* in 1906 and *Shigella boydii* in 1921 (REFS 15,16). *Shigella* bacteria are non-motile, non-sporulating and non- or late-lactose-fermenting, and classical taxonomy places all *Shigella* spp. into one major group, which is distantly related to *E. coli*[17]. However, even on their initial characterization, the biochemical and morphological proximity of members of the genus *Shigella* with *E. coli* was noted[13]. Biochemical differences exist between the two genera: *S. dysenteriae* is negative in an indole reaction and cannot ferment mannitol[13], and all *Shigella* spp. are negative for lysine decarboxylation (LDC)[18], whereas the opposites are true for *E. coli*. Current serological classification divides the genus *Shigella* into four species (also known as subgroups), which are further subdivided into serotypes according to type-specific antigens: *S. dysenteriae* (subgroup A) has 15 serotypes; *S. flexneri* (subgroup B) has 19 serotypes and subserotypes; *S. boydii* (subgroup C) has 20 serotypes; and *S. sonnei* (subgroup D) consists of a single serotype.

S. flexneri is currently the major cause of bacillary dysentery in low-income settings (in parts of Asia and sub-Saharan Africa, this species accounts for up to 62% of all *Shigella* spp. infections), whereas *S. sonnei* is the most common pathogen in transitional or high-income countries (up to 80% of all *Shigella* spp. infections in Europe and North America are caused by this species)[19]. *S. boydii* and *S. dysenteriae* cause <5% each of all cases of shigellosis globally. Notably, *S. dysenteriae* was the main cause of dysentery when it was first identified more than a century ago, but today it is infrequently isolated from patients with dysenteric diarrhoea[7,19]. It is thought that poor sanitation, malnutrition and unavailability of clean water, and an exceptionally low infectious dose, genomic plasticity and an ability to accept antimicrobial-resistance genes are all potential reasons why *Shigella* spp. are such successful pathogens and why particular human populations are specifically vulnerable to infection with these species.

The genomics revolution has revealed the dynamic genome plasticity of *Shigella* spp. and their close evolutionary history with *E. coli*[20]. The pathogenesis of *Shigella* spp. depends on a large virulence plasmid that, during its enigmatic evolutionary history, has acquired several factors that are essential for invasion and subversion of host defences[21]. Recent advances in high-throughput genomics and phylogenetics have detailed the emergence and spread of different *Shigella* spp. serogroups, and this information can in turn be used to inform control and public health polices for shigellosis[22–24].

In this Review, we discuss the evolution of *Shigella* spp. to highly specialized, human-specific pathogens, taking into account both insights from traditional genotyping methods and current perspectives achieved from phylogenomics. We focus on the most commonly isolated *Shigella* spp., *S. sonnei* and *S. flexneri*, as these are the dominant species responsible for the global burden of shigellosis.

The evolutionary history of *Shigella* spp.

The acquisition of the *Shigella* virulence plasmid was the key event in the formation of the different *Shigella* spp.[25], but the origins of this plasmid and the relationship between the species was contentious for a long time. The advent of DNA sequencing and accompanying phylogenetic analyses have led to a much clearer picture of the evolutionary relationships between the different *Shigella* spp. and their emergence from *E. coli*.

The phylogenetic relationships of* Shigella *spp. Pioneering research in the early genomics era, carried out by aligning and comparing the DNA sequences of eight chromosomal housekeeping genes, found the *Shigella* genus to contain three major clades or clusters (C1, C2 and C3) and a limited number of outliers, all of which are distinct from, but nested within lineages of *E. coli*[26]. A further examination of 23 chromosomal genes reached a similar phylogenetic conclusion, albeit with increased resolution, subdividing C1 into 3 subclusters (SC1, SC2 and SC3)[27] (FIG. 1a). Most *Shigella* spp. serotypes are distributed over the three major clusters, demonstrating an incongruence between evolutionary history and the conventional serology-based nomenclature. Cluster C1 contains a combination of serotypes from *S. dysenteriae* and *S. boydii*, as well as *S. flexneri* serotype 6: SC1 includes only *S. dysenteriae* (serotypes 3, 4, 6, 9, 11, 12 and 13); SC2 contains mostly *S. boydii* (serotypes 1, 3, 6, 8, 10 and 18), as well as *S. dysenteriae* serotype 5; and SC3 is composed of three other *S. boydii* serotypes (2, 4 and 14) and *S. flexneri* serotype 6. Cluster C2 comprises *S. boydii* (serotypes 5, 7, 9, 11, 15, 16 and 17) and *S. dysenteriae* serotype 2. All *S. flexneri* serotypes except 6 (that is, 1, 2, 3, 4, 5, X and Y) fall into cluster C3, as well as *S. boydii* serotype 12. In this analysis, C2 and C3 were found to share a more recent common ancestor than their common ancestor with C1, thus emphasizing the close phylogenetic relationship between these two clusters.

An analysis of short DNA sequences yielded an estimation of the age of the various clusters (50,000–270,000 years for each of C1 and C2; 35,000–170,000 years for C3)[26]; however, whole-genome sequencing and Bayesian phylogenetic tools are expected to provide a more accurate genome-wide dating of these clusters. Notable outliers not belonging to any of the three major clusters include *S. sonnei*, *S. dysenteriae* 1, 8 and 10, and *S. boydii* 13 (FIG. 1a). The position of *S. boydii* 13 on the tree topology indicates that it is also distant from the *E. coli*–*Shigella* clade. This genetic distance is consistent with the finding that *S. boydii* 13 and an *Escherichia albertii* group form a discrete lineage that separated from an *E. coli* ancestor ~28 million years ago[28].

As highlighted above, *S. sonnei* is an outlier from the other *Shigella* spp., and the precise phylogenetic relationship between *S. sonnei* and the other *Shigella* spp. remains ambiguous. It is assumed that *S. sonnei* emerged more recently than the other *Shigella* spp. and serotypes[29].

Disability adjusted life years (DALY). A measure of overall disease burden, expressed as the cumulative number of years lost owing to ill health, disability or early death.

Pathovars
Groups of bacterial strains that have similar characteristics and are differentiated at the subspecies level on the basis of their distinctive pathogenicity in one or more hosts.

Lysine decarboxylation (LDC). A reaction that is used in a biochemical test to determine the ability of a microorganism to use lysine as a source of carbon for growth. In a positive LDC test, lysine is metabolized into the amine cadaverine through the activity of the enzyme lysine decarboxylase.

© 2016 Macmillan Publishers Limited. All rights reserved.

Microfold cell
(M cell). A specialized epithelial cell type found in the follicle-associated epithelium of the gastrointestinal tract. Their function is to transport macromolecules and microorganisms across the epithelial barriers to the immune cells, thus inducing mucosal immunity.

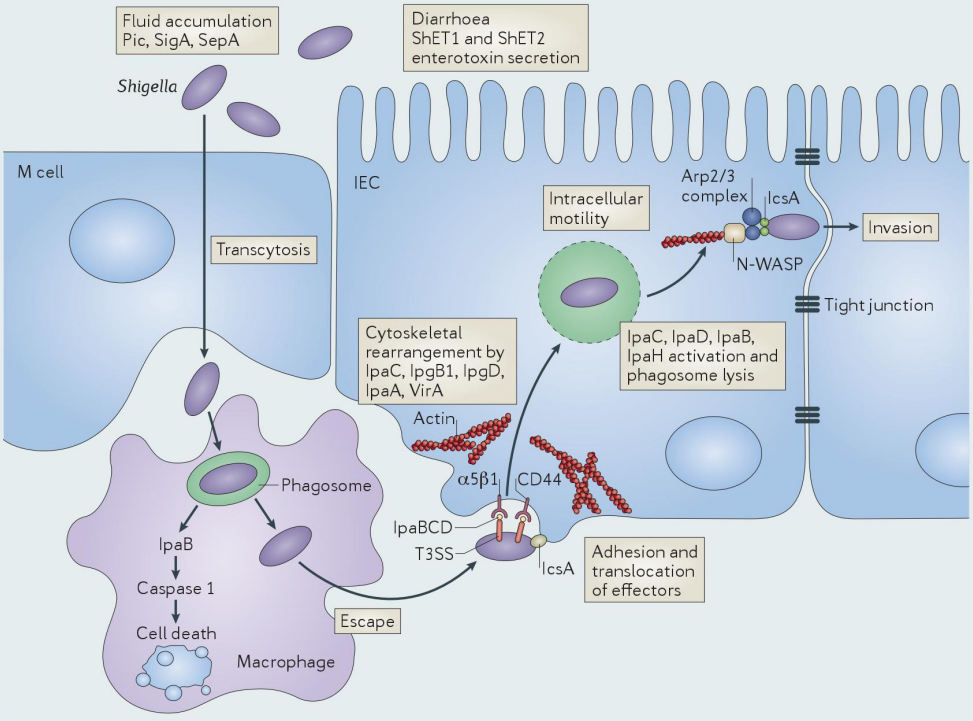
Transcytosis
The selective vesicular transport of macromolecules from one side of the cell to the other while maintaining the unique compositions of these vesicular environments.

Box 1 | The molecular mechanisms of *Shigella* spp. pathogenesis

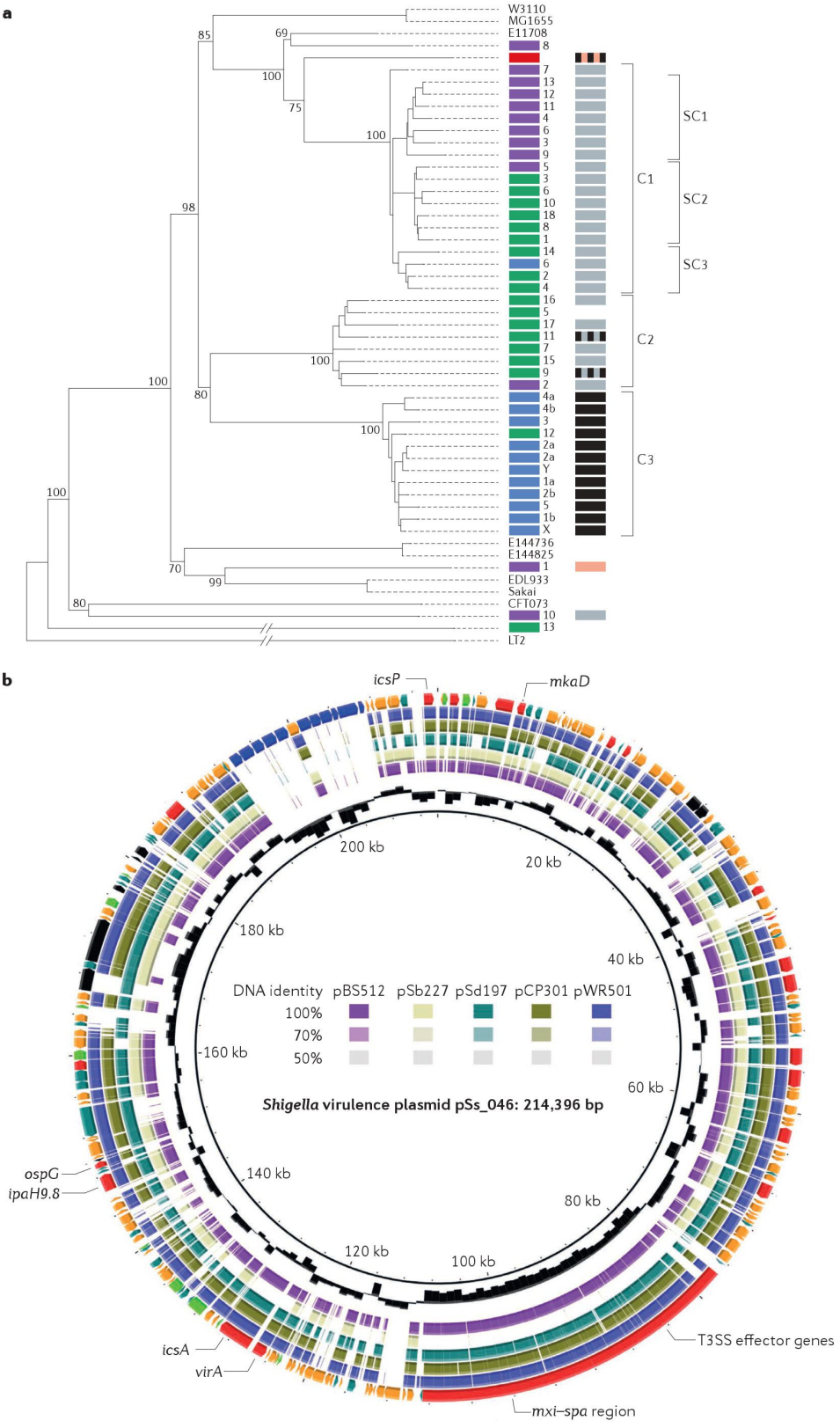
Infection with *Shigella* spp. usually results in self-limiting diarrhoea, which is initially watery and then bloody and/or mucus containing. Pathogenesis is tightly orchestrated by the Mxi–Spa type III secretion system (T3SS) and a plethora of effector proteins encoded on the virulence plasmid, including invasion plasmid antigen B (IpaB), IpaC and serine protease SepA, and has been proposed to be a multistep process[37,125]. First, the *Shigella* sp. crosses the intestinal epithelium via a microfold cell (M cell)[126]. This process, termed transcytosis, allows the *Shigella* sp. to cross the physical protective barrier of the host and exposes the bacterium to macrophages in the submucosa of the gastrointestinal tract. The bacterium is phagocytosed, but escapes destruction by inducing rapid macrophage apoptosis through a caspase 1-dependent pathway[127,128]. After its release, the bacterium comes into contact with the basolateral side of an intestinal epithelial cell (IEC) and triggers effector-mediated endocytosis. The Mxi–Spa T3SS needle-like structure is coated by the protein IpaB at the tip, and IpaB has a high affinity for cholesterol-rich areas of eukaryotic cell membranes[129]. The adhesion of the *Shigella* sp. to an IEC is achieved by the binding of IpaB and IpaBCD to the host hyaluronan receptor CD44 and α5β1 integrin, respectively[130,131]. Recently, IcsA was shown to function as an indispensible adhesin to promote contact with host IECs[119]. This close contact facilitates the fusion of the T3SS apparatus and the host cell membrane, leading to the translocation of further effectors. Successful engulfment requires extensive reorganization of the host cell cytoskeleton and modulation of other intracellular pathways, involving the activity of the bacterial effectors IpaC, IpgB1, IpgD, IpaA and VirA in concert[132–136]. *Shigella* spp. are also capable of invading the IECs on the apical side through filopodial capture, which probably requires an interaction between IpaB, IpaD and receptors present on the filopodial extensions but without constitutive binding[137]. After invasion, the *Shigella* sp. quickly lyses its surrounding phagosome by activating IpaC, IpaD, IpaB and IpaH, of which IpaC has been shown to have the pivotal role[138,139]. The cytoplasm serves as a new niche for *Shigella* sp. survival and replication, until diminished resources prompt the invasion of neighbouring cells. The motility of the intracellular *Shigella* sp. is supported by IcsA through the recruitment of the host's neural Wiskott–Aldrich syndrome protein (N-WASP) and actin-related protein 2/3 (ARP2/3) complex, creating a nucleator site for directional actin polymerization to propel the bacterium through the cytoplasm[140]. Intercellular dissemination occurs by this same mechanism when the *Shigella* sp. is endocytosed primarily at tricellular tight junctions, and a new cycle of release–replication–spread takes place.

In contrast to the rapid killing of macrophages, *Shigella* sp. must prolong the survival of infected IECs to ensure its own replication. Inside IECs, the secreted IpaB targets mitotic arrest deficient 2-like protein 2 (MAD2L2; also known as MAD2B), arresting cell maturation[141]. Inhibition of cell death and detachment are triggered by the activity of IpgD and OspE, respectively[142,143]. Cellular secretory functions, including that of pro-inflammatory cytokines, are blocked by the disruption of the Golgi and ER–Golgi trafficking by the activation of the multifunctional effectors IpaB and VirA, respectively[144,145].

Profuse diarrhoea in a *Shigella* sp. infection is attributed to the production of *Shigella* enterotoxin 1 (ShET1) and ShET2, which are encoded in the genomic pathogenicity island *Shigella* island 1 (SHI-1) and the virulence plasmid pINV, respectively[146,147]. In addition, several members of the serine protease autotransporter of Enterobacteriaceae (SPATE) family enhance bacterial virulence. The chromosomally encoded Pic and SigA, as well as the pINV-encoded SepA, intensify intestinal fluid accumulation in an animal model, a hallmark of the early stages of the infection[42,148].



© 2016 Macmillan Publishers Limited. All rights reserved.



© 2016 Macmillan Publishers Limited. All rights reserved.

O antigen
A repeating glycan polymer attached to the outer core in lipopolysaccharide. This structure is on the very outer surface of the bacterial cell and is therefore a target for recognition by the host immune system.

Unlike the other *Shigella* spp., *S. sonnei* expresses an O antigen, encoded by a genetic locus that is also found in the genetically distant Gram-negative organism, *Plesiomonas shigelloides*[29]. A sequence comparison of the O antigen loci from *S. sonnei* and *P. shigelloides* predicts that the O antigen genes diverged approximately 10,000 years ago, placing an upper limit on the age of the formation of *S. sonnei*[29]. However, a more recent study using whole-genome sequencing data from globally distributed isolates estimated that all extant strains of *S. sonnei* descend from a common ancestor that existed <400 years ago, implying that a historical evolutionary bottleneck might have resulted in the extinction of the pre-existing *S. sonnei* strains[22].

The early sequence-based genotyping studies described above largely resolved the phylogenetic relationships of the different *Shigella* spp., but more recent studies have exploited *Shigella* spp. and *E. coli* whole-genome sequences to investigate the evolutionary relationship between these two taxa in more detail. Phylogenetic trees for the entire *E. coli–Shigella* group were constructed using an alignment-free feature frequency profile (FFP), which compares genomes based on the frequencies of oligonucleotide sequences with an optimal length for analysis (so-called features)[30]. These phylogenies, together with those deduced from other studies using core genetic features (present in all genomes and with low variability), have confirmed that the genus *Shigella* is composed of several clusters interspersed in the *E. coli–Shigella* phylogeny, strongly supporting the notion that *Shigella* spp. have emerged from several *E. coli* ancestors on multiple independent occasions[31–33]. The phylogenomic structure of the genus *Shigella* derived from a collection of 336 *E. coli–Shigella* isolates correlates with the grouping from the aforementioned studies based on a limited number of genetic markers[34]. In addition, whole-genome resolution phylogenomics also resolves the context for the origins of these major clades: it places C1 and *S. sonnei* in *E. coli* group B1; C2 and C3 in *E. coli* group A; and *S. dysenteriae* 1 in *E. coli* group E[34]. This supports the theory that the phenotypic similarity observed across the *Shigella* spp. is the result of convergent evolution, in which different *Shigella* founders independently gained genes that facilitate invasive pathogenicity. Only one *E. coli* pathovar, EIEC, has also acquired invasiveness; EIEC comprises several discrete lineages and exhibits pathogenic and biochemical features that are indistinguishable from those of *Shigella* spp. Notably, both EIEC and *Shigella* spp. harbour an analogous virulence plasmid, are non-motile and show a negative LDC test[35]. These similarities have led to the speculation that EIEC represents a distinct non-toxin-producing *Shigella* 'prototype', which could be a precursor for a 'complete' *Shigella* sp. if selective pressure favours further adaptation of this invasive *E. coli* pathovar[36].

◀ Figure 1 | **The phylogenetic structure of the four *Shigella* spp. and the signature virulence plasmid. a** | A neighbour-joining phylogenetic tree generated by sequencing 23 chromosomal genes[27]. Strains are labelled by serotype and coloured by species: *Shigella sonnei* in red, *Shigella flexneri* in blue, *Shigella boydii* in green and *Shigella dysenteriae* in purple; *Escherichia coli* isolates are uncoloured. Bootstrap values of >50% are indicated at the major nodes, and the three major *Shigella* genus clusters (C) and subclusters (SC) are indicated. The carriage of the two specific isoforms of virulence plasmids is additionally indicated in the second column of coloured blocks: pINV A (grey), pINV B (black), either pINV A or pINV B (hatched black and grey), a unique form of pINV (pink), and either pINV B or a unique isoform (hatched black and pink). **b** | A comparative gene map of the *Shigella* virulence plasmid, using the *S. sonnei* virulence plasmid pSs_046 as a reference; the innermost ring represents pSs_046, with coordinates. The second ring (black) shows the GC content of the reference pSs_046 sequence. The following purple, pale green, teal, khaki, and blue rings show BLASTN comparisons between pSs_046 and the virulence plasmids of *S. boydii* str. BS512, *S. boydii* str. Sb227, *S. dysenteriae* str. Sd197, *S. flexneri* F2a str. 301 (pCP301) and *S. flexneri* F5a (pWR501), respectively. The outer ring represents annotations of genes or genetic clusters based on function: known virulence factor genes (red); plasmid replication, transfer and maintenance genes (black); transposon, phage-borne and insertion sequence elements (orange); genes encoding hypothetical proteins (teal); the *S. sonnei*-specific O antigen biosynthesis cluster (blue); and genes encoding proteins with other known functions (green). *ipa*, invasion plasmid antigen gene; *icsP*, also known as *sopA*; T3SS, type III secretion system. Part **a** is modified with permission from REF. 27, Springer.

The* Shigella *virulence plasmid. The *Shigella* virulence plasmid, which can be as large as ~220 kb, encodes essential virulence factors that facilitate the invasion and spread of *Shigella* spp. into human macrophages and enterocytes[37] (BOX 1). The virulence plasmid contains the conserved 30 kb *mxi–spa* locus, which encodes the Mxi–Spa type III secretion system (T3SS), and genes encoding invasion plasmid antigens (Ipas). The Mxi–Spa T3SS is a molecular syringe that injects effector proteins directly into host cells. This secretion apparatus enables a complex interaction between the bacterium and the host cell, ultimately resulting in a disruption of the intestinal mucosa and the distinctive symptoms of bacillary dysentery. Therefore, the virulence plasmid is the key molecular signature of *Shigella* spp. pathogenesis and is fundamental for initiating infection and manipulating the immune response of the host (BOX 2).

Various DNA sequencing projects have been carried out across several different *Shigella* spp. lineages to elucidate the structure and functions of the virulence plasmid. These projects have uncovered a complex plasmid configuration with a mosaic nature, which is the result of numerous horizontal gene transfer and rearrangement events[21,38,39] (FIG. 1b). The evaluation of three genes in the *mxi–spa* region (*mxiA*, *mxiC* and *ipgD*) revealed two isoforms of the *Shigella* virulence plasmid (pINV A and pINV B) with greater divergence in *ipgD* than in the two *mxi* genes[25]. pINV A and pINV B exhibited incompatibility grouping (plasmids of the same incompatibility group cannot be stably inherited in the same cell)[40]. When plasmid subtype is mapped onto the *Shigella* spp. phylogeny (FIG. 1a), all C1 isolates harbour pINV A, whereas all C3 isolates possess the pINV B isoform. Both forms of the plasmid can be found in C2 isolates. The outlier strains harbour either of the two plasmid forms, which is a sign of lateral gene transfer in their history. For example, *S. dysenteriae* 10 and most EIEC strains harbour pINV A, whereas *S. sonnei* retains pINV B[36,41]. By contrast, *S. dysenteriae* 1 harbours a unique mixed plasmid form (*ipgD* derived from pINV A, and *mxiA* and *mxiC* derived from pINV B). This suggests that several ancestral virulence plasmids, from an unknown source, have entered into a diverse background of *E. coli*

© 2016 Macmillan Publishers Limited. All rights reserved.

Box 2 | **The immune response against *Shigella* spp.**

Immune modulation has a major role in *Shigella* spp. pathogenesis, beginning with the ability of the organisms to manipulate the innate immune response[149]. In the initial stages of infection, rapid killing of infected macrophages by the caspase 1 pathway releases the pro-inflammatory cytokines interleukin-1β (IL-1β) and IL-18 (REF. 150). This acute inflammation, heightened by the secretion of CXC-chemokine ligand 8 (CXCL8; also known as IL-8) from infected epithelial cells, triggers the transepithelial migration of neutrophils and an influx of more *Shigella* sp. cells[149,151]. By contrast, inside enterocytes, the *Shigella* sp. releases a cascade of effectors, such as MkaD, which inactivate the mitogen-activated protein kinases (MAPKs) p38 and ERK2; OspG, which targets the ubiquitin-conjugating enzyme E2; OspI, which deamidates the E2 enzyme UBC13 (also known as UBE2N); OspZ, which prevents the nuclear translocation of the transcription factor NF-κB; and invasion plasmid antigen H9.8 (IpaH9.8), which targets the NF-κB essential modulator (NEMO; also known as IKKγ) complex. These pathways inhibit the NF-κB-dependent inflammatory responses, masking the bacteria from detection by the immune system and maintaining their intracellular proliferation[152–154].

The death of B cells and T cells has been observed during infection with *Shigella flexneri*[155]. The type III secretion system (T3SS) effector IpgD has been shown to impair the migration of activated T cells *in vitro* probably through phosphatidylinositol hydrolysis, which impedes the reorganization of the cytoskeleton[156]. Inhibition of T cell migration compromises T cell contact with antigen-presenting cells and thus dampens the adaptive immune response. An *in vivo* study of *Shigella flexneri* infection in lymph nodes has confirmed the capacity of the bacterium to invade T cells and arrest their movements by the T3SS[157]. In addition, there is evidence that the T3SS-coating protein IpaD targets Toll-like receptor 2 (TLR2) on B cells and induces apoptosis, irrespective of invasion[158].

founder strains. Such introductions include pINV A and pINV B into major *Shigella* clusters C1 and C3, respectively, thus giving rise to these two lineages. Independent acquisitions of either plasmid isoform by *Shigella* spp. isolates not belonging to the main clades, as well as lateral gene transfer in the C2 isolates, complicate the evolutionary history of *Shigella* spp.

Plasmid sequences have also been compared for five virulence genes (*mkaD* (also known as *ospF*), CP0014, *parA*, *parB* and *repA*) located outside the core entry region (defined as the ~30 kb cluster encoding the T3SS and associated effector proteins that facilitate the mechanistic invasion of the bacteria into enterocytes). The constructed phylogeny was consistent with the one based on chromosomal genes[27], except for a close relationship between C1 and C2 isolates in the plasmid phylogeny; by contrast, the C2 and C3 clusters showed close proximity in the chromosomal phylogeny. The authors of the plasmid phylogeny report argued that the virulence plasmid acquired by C1 and C2 isolates differed from the one obtained by C3 isolates. Interestingly, the virulence plasmids from the outliers *S. dysenteriae* 1 and *S. sonnei* share considerable homogeneity and can be grouped together outside of the three major clusters. These data suggest that *Shigella* spp. have arisen on several independent occasions owing to the transmission of multiple virulence plasmid forms to many *E. coli* ancestors. The authors suggested that the subsequent loss of the *tra* locus, which aids the exchange of plasmids between bacteria by conjugation, on the virulence plasmid restricted its transmissibility and enabled parallel evolution of the virulence plasmid and the bacterial chromosomes, thus creating the several discrete *Shigella* lineages observed today.

Gain and loss of gene function. The ability to invade host cells and escape the competitive environment of the gastrointestinal tract was pivotal in the emergence of *Shigella* spp. Although acquisition of the virulence plasmid is a 'foothold moment' in the evolution of this pathogen, it is not the only long-term evolutionary change. Numerous other plasmids with different functions have been crucial during the evolution of *Shigella* spp. (BOX 3). In addition to the genes encoded on the virulence plasmid, several clusters of horizontally acquired genetic material, carrying genes that facilitate interactions with the host and contribute to pathogenesis, have been incorporated into the chromosome of *Shigella* spp.

These pathogenesis-associated genomic regions are pathogenicity islands (PAIs) and have various functions; the largest PAI encodes an enterotoxin (*Shigella* island 1 (SHI-1)) and it enables the sequestration of iron (SHI-2, SHI-3 and *sitABCD*), the ability to modify the O antigen (SHI-O) and resistance to antimicrobials (*Shigella* resistance locus (SRL))[37,42–46]. PAIs have enhanced the virulence and adaptability of *Shigella* spp. and are commonly associated with bacteriophage integrases, which highlights the fact that bacteriophages had a major role in the evolution of *Shigella* spp. One such bacteriophage-associated element is the Shiga toxin (Stx) prophage in *S. dysenteriae* 1; Stx expression can have severe complications, including haemolytic uraemic syndrome (HUS). Recently, an alternative prophage (φPOC-J13) encoding Stx1a was identified in several clinical isolates of *S. flexneri* and *S. dysenteriae* 4 from patients returning from or residing in Hispaniola[47–50]. Unlike the cryptic prophage in *S. dysenteriae* 1, φPOC-J13 seems to be capable of disseminating the stx_{1a} gene into other *Shigella* spp. isolates by transduction[50]. Insertion sequence elements — small transposable DNA sequences that can 'jump' within bacterial genomes — are also highly abundant in *Shigella* spp. chromosomes and virulence plasmids. These elements have shaped the genome architecture of *Shigella* spp., causing gene inactivation and genome rearrangement[20,21,51]. An analysis of >400 genomes from a range of bacterial species found that, in relation to genome size, *E. coli* and *Shigella* spp. possess the highest number of insertion sequence elements[52].

Linked to this insertion sequence expansion, *Shigella* spp. genomes have also undergone substantial functional gene loss[53]. Similar phenomena have been observed in other human-restricted pathogens, such as *Yersinia pestis*, *Mycobacterium leprae* and *Salmonella enterica* subsp. *enterica* serovar Typhi[54–56]. The modes of gene inactivation are variable in different *Shigella* spp. strains and range from the complete deletion of a locus, to missense point mutations, to insertions. However, gene inactivation has occurred preferentially in specific genetic regions and operons rather than being randomly distributed throughout the genome[20,57]. Independent inactivation of the same or functionally similar genes in different *Shigella* spp. represents a major pathway of convergent evolution, resulting in similar phenotypic changes that are associated with adaptation to new niches. For example, different mutations have resulted

Conjugation
The transfer of genetic material between bacterial cells through cell-to-cell contact or by a bridge-like connection between two cells.

Transduction
A mode of horizontal gene transfer whereby genetic material is transferred from one bacterium to another by a virus.

Flagellum
A multiprotein thread-like structure protruding from prokaryotic or eukaryotic cells that is used for motility and for the sensory perception of extracellular chemicals and temperature.

Fimbriae
Appendages composed of the protein curlin and found on many Gram-negative and some Gram-positive bacteria. Fimbriae are used mainly for adherence to bacterial cells, host cells and abiotic surfaces.

© 2016 Macmillan Publishers Limited. All rights reserved.

Box 3 | ***Shigella* spp. plasmids**

The main plasmid that enables the intracellular lifestyle of *Shigella* spp. is the large virulence plasmid pINV, which encodes a type III secretion system (T3SS) and an arsenal of virulence factors[21,159]. In *Shigella sonnei*, pINV also harbours an O antigen biosynthesis cluster homologous to that of *Plesiomonas shigelloides*[29]. Under laboratory conditions, this plasmid is usually lost from *S. sonnei* after subsequent culturing, a factor that has undergone limited detailed investigation[22,160]. An additional small plasmid, pHS-2, harbours a single gene, chain length determinant (*cld*; also known as wzz_{pHS-2}), which is instrumental in controlling O antigen chain length and, thus, determining resistance to serum killing[116,161]. Although pHS-2 was originally thought to be associated with reactive arthritis induced by shigellosis as a result of a hypothesized molecular mimicry of human HLA-B27 by a pHS-2-encoded protein, this suggestion has been challenged by recent findings, including the occurrence of reactive arthritis in patients infected with pHS-2-negative *S. sonnei*[162–164]. Colicins are bactericidal proteins that act on those bacterial species which are closely related to the colicin-producing bacteria, and it has been proposed that colicins provide *Shigella* spp. with a competitive advantage against other susceptible *E. coli* or *Shigella* spp. populations[165]. It has been speculated that the introduction of colicin-encoding plasmids initiated the recent clonal expansions of specific *S. sonnei* populations, such as the acquisition of pDPT1 by *S. sonnei* in Vietnam and of pSSE3 by *S. sonnei* in South Asia[75,165,166].

Shigella spp. have also acquired several antimicrobial-resistance plasmids. The small 8 kb spA plasmid, encoding resistance to sulfonamide, streptomycin and tetracycline, was introduced into multiple *S. sonnei* populations during the latter part of the twentieth century, rendering treatments with these drugs ineffective[22]. More recently, stably inherited extended-spectrum β-lactamase (ESBL)-producing plasmids are increasingly being recovered from both *S. sonnei* and *S. flexneri* in developing countries, which raises concerns about restricted therapeutic options, in particular as fluoroquinolone resistance is increasing worldwide[19,24,75]. Other modern antimicrobials are also at risk of becoming inactive, as plasmid pKSR100, carrying resistance against azithromycin, erythromycin, β-lactams and aminoglycosides, has been commonly found in the *S. flexneri* 3a lineage circulating in the men who have sex with men (MSM) community[101].

in the loss of flagellum biosynthesis and specific structures of the fimbriae[20,58]. Importantly, flagellum loss results in reduced immunogenicity and in evasion of the human immune system, as flagellin is a pathogen-associated molecular pattern (PAMP)[59]. In addition, computational reconstruction of metabolic functions (or their loss) based on genomic data groups the different *Shigella* spp. together and away from *E. coli* pathovars and commensals, purely by their catabolic function[60]. For example, the *E. coli* genes *cadA* and *nadAB* (responsible for the synthesis of cadaverine and the NAD precursor quinolinate, respectively) hinder intercellular spread, phagosomal escape and antigen secretion[61–63]. Likewise, *ompT* and *argT* inhibit intracellular motility and invasive capacity, respectively[64,65]. Therefore, the loss of these genes in *Shigella* spp. ensures patho-adaptation for an intracellular lifestyle. Alternatively, the loss of gene function can increase the survival rate in a new niche. Disruption of *speG*, which encodes spermidine acetyltransferase, leads to the accumulation of the polyamine spermidine, which acts as scavenger of free radicals and thereby provides resistance to oxidative stress in macrophages[66].

In comparison to *E. coli*, *Shigella* spp. have lost more genes, which is attributed to a reduction in genome-wide purifying selection and the fixation of inactivating mutations without greatly compromising fitness. It has been suggested that this resulted from a decrease in effective population size when *Shigella* spp. became human-restricted pathogens, compared with their *E. coli* ancestor. In addition, the intracellular niche in which *Shigella* spp. began to thrive imposed a more relaxed selective pressure owing to abundant resources and a relative lack of competitors[67,68].

Pathogen-associated molecular pattern
(PAMP). A set of specific molecules that are present on groups of pathogens and are recognized by the innate immune system. These conserved molecular motifs in bacteria, such as lipopolysaccharide and flagellin, are usually recognized by Toll-like receptors and other pattern-recognition receptors.

Cadaverine
A diamine with a putrid odour. It is the product of lysine decarboxylation.

Quinolinate
A dicarboxylic acid generated as the downstream product of tryptophan catabolism. It acts as a substrate for the biosynthesis of nicotinic acid mononucleotide and, ultimately, the formation of the coenzyme NAD.

Genomic insights into *Shigella* spp.

The advent of high-throughput whole-genome sequencing has permitted the detection of genomic variation in the form of single-nucleotide polymorphisms (SNPs) and accessory genome content. These techniques give us an unprecedented view of how *Shigella* spp. have emerged and been transmitted globally, and how antimicrobial resistance has swept through the population throughout the later part of the twentieth century.

Shigella sonnei. Although *S. sonnei* is the most common *Shigella* species in middle-income and high-income countries, the recent emergence of this species in transitional lower-income countries has highlighted the need for more effective surveillance systems and has opened new avenues of vaccine research[69–73] (BOX 4). In a key study, the genomes of 132 globally representative *S. sonnei* isolates were sequenced and analysed to investigate the recent evolution of the species[22]. Three main lineages of *S. sonnei* were identified (I, II and III), which share a most recent common ancestor <400 years ago. All lineages probably originated in Europe, as the oldest lineages, and the majority of all genetic diversity, were detected in European isolates (FIG. 2a). Although all three lineages were distributed in Europe, not all of them have spread globally. Isolates from Asia, Africa and South America were predominantly representatives of the more recently expanded lineage III. Lineage III (particularly clade Global III) emerged in the 1970s and spread internationally in the 1980s and 1990s, establishing more distant endemic populations in other regions of the world (FIG. 2a). Importantly, there was a correlation between global dissemination and the acquisition of resistance to multiple antimicrobials. Resistance was mediated through the gain of class II integrons and mutations in DNA gyrase subunit A (*gyrA*), which encodes the target protein for fluoroquinolones (a family of broad-spectrum antimicrobials). These modifications probably resulted from strong selective pressures induced by antimicrobial exposure; indeed, antimicrobial resistance may be advantageous in promoting post-symptomatic shedding of bacteria and sustained short-term transmission in the host population[74].

A study that investigated more than 250 *S. sonnei* samples in Vietnam expanded the observations of the global study[75]. Genomic and phylogeographical analyses showed that the Global III lineage became established in Ho Chi Minh City following the reunification of the country in 1975. The founder clone later spread north to other provinces, where it established, albeit after multiple introductions, further discrete endemic populations (FIG. 2b). Clonal expansion in these regions contributed to the increase in *S. sonnei*-associated dysentery in Vietnam. In larger human populations, such as in Ho Chi Minh City, a series of bottlenecks in the bacterial population and the stepwise accumulation of antimicrobial resistance were observed, probably as a consequence of the

© 2016 Macmillan Publishers Limited. All rights reserved.

Box 4 | **Vaccines against *Shigella* spp.**

The adaptive immune response to *Shigella* spp. infection largely targets the bacterial O antigen[167], rendering this structure a sound candidate for vaccine development. However, this approach is hindered by the wide geographical distribution of numerous serotypes, highlighting the requirement for large-scale surveillances. It was calculated that ~64% of the shigellosis episodes in the Global Enteric Multicenter Study (GEMS) project were caused by only four serotypes, *Shigella sonnei*, *Shigella flexneri* 2a, *S. flexneri* 3a and *S. flexneri* 6 (REF. 113). Owing to the extensive cross-protection provided by *S. flexneri* 2a and *S. flexneri* 3a O antigens against other serotypes of this species[168], a quadrivalent vaccine composed of the O antigens of the four serotypes could in theory provide ~88% coverage (in the case of a 100% efficacious vaccine) against shigellosis[113].

The rise of *S. sonnei* in economically transitioning nations poses questions about the management and control of shigellosis worldwide and highlights the feasibility of a vaccine against a single-serotype enteric pathogen. It has been suggested that because *S. sonnei* and *Plesiomonas shigelloides* have an almost identical O antigen structure, exposure to the latter in contaminated water provides immunity to the former by passive 'environmental' immunization[29]. An improvement in access to clean water facilities in transitional economies reduces the occurrence of environmental *P. shigelloides* and any potential cross-protective passive immunization. This may explain why *S. sonnei* is able to thrive in transitional countries. However, despite being a fascinating hypothesis, the exact correlation between these two phenomena remains unclear[169].

Shigella vaccine development involves consortia of experts, and it has been reviewed elsewhere[170–172]. Currently available candidate vaccines can be classified into three major approaches: those targeting specific O antigens, those targeting common conserved proteins and those targeting a combination of both. Live-attenuated variants of *S. sonnei*, *S. flexneri* 2a and *Shigella dysenteriae* 1 are entering different phases in clinical studies; these variants have been engineered to harbour mutations in essential virulence genes, such as *guaA*, *guaB*, *icsA*, or enterotoxin genes *senA* (*Shigella* enterotoxin 2), *senB*, *stxA* (Shiga toxin subunit A) or *stxB*. Furthermore, serotype-specific lipopolysaccharides conjugated with carrier proteins (*Pseudomonas* exoprotein A or tetanus toxoid) are also potential candidates. Purified virulence plasmid-encoded proteins invasion plasmid antigen B (IpaB) and IpaD were shown to confer protection in animal models, as well as the IcsP and SigA proteins[173]. Invaplex, a combination of highly conserved *Shigella* spp. IpaBCD and lipopolysaccharide, induces a serotype-specific immune response after intranasal delivery[174].

strong selective pressure exerted by the high use of antimicrobials in the country. Furthermore, plasmid pDPT1, encoding an E5 type colicin (a bactericidal toxin with RNA degradation potency) and an associated immunity protein (protecting the producer from the activity of the corresponding colicin), became fixed in the Ho Chi Minh City population following the first selective sweep in 1994, providing a crucial selective advantage over other non-immune *Shigella* spp. and *E. coli* strains. In the 2006 selective sweep, the population acquired plasmid pKHSB1, which harbours an extended-spectrum β-lactamase (ESBL) gene. This explains the sudden increase in the isolation rate of cephalosporin-resistant *S. sonnei* in the following years in the region[76]. The acquisition of a plasmid conferring resistance to third-generation cephalosporins (FIG. 2b) reoccurred in satellite populations in the central region of Vietnam, namely in the Khanh Hoa province. Similarly, other signs of convergent evolution included the independent emergences of *gyrA* mutations in Ho Chi Minh City and other provinces, reducing the susceptibility to fluoroquinolones. With such a detailed understanding of the *S. sonnei* population in Vietnam, the authors suggested that *S. sonnei* could act as a sentinel organism for the surveillance of human enteric bacterial pathogens by providing a tractable window onto the circulating antimicrobial-resistance elements in other Gram-negative enteric bacteria in a specific region. Indeed, the transfer of third-generation cephalosporin resistance plasmids between *S. sonnei* and commensal *E. coli* in the human gut might occur, as the expansion of the *S. sonnei* population during an episode of infection greatly increases the chance of contact between these two organisms[77].

Class II integrons
Mobile genetic elements that are capable of carrying genes, including antimicrobial-resistance genes, and integrating into bacterial chromosomes by site-specific recombination. An integron contains at least an integrase, an attachment site and a promoter. Classification is based on the type of integrase.

Shigella flexneri. Alongside *S. sonnei*, *S. flexneri* remains a major aetiological agent of bacillary dysentery, particularly in low-income and middle-income countries. Much of our epidemiological knowledge about *S. flexneri* comes from serotyping data. *S. flexneri* serotypes differ in their O antigens, and there is experimental evidence that the O antigen conformation is important for invasion and the evasion of innate immunity[78]. However, serotype conversion (that is, the modification of the serotype in a clonal population) is well documented in *S. flexneri* and mediated by bacteriophages and plasmids carrying genes that contribute to variation of the O antigen structure. The bacteriophages often integrate as prophages into the chromosomal *thrW* tRNA site, for prophages carrying the glycosylation (*gtr*) operon, or into the *argW* tRNA site, for those carrying the *O*-acetylation (*oac*) gene, and lead to changes in the O antigen structure[79,80]. Many O antigen-modifying bacteriophages have been identified to date, including SfI, SfII, Sf6, SfIV, SfV and SfX, which convert *S. flexneri* Y into serotypes 1a, 2a, 3b, 4a, 5a and X, respectively[81–86]. Furthermore, several novel *S. flexneri* serotypes have been discovered in the past decade, which complicates the epidemiology and potential protective efficacy of any potential O antigen-based vaccines (BOX 4).

The emergence of novel *S. flexneri* serotypes has been widely observed. For example, *S. flexneri* 1c was first identified in Bangladesh in the 1980s, and an unrelated clone of this serotype was then also found to be prevalent in rural northern Vietnam and several other Asian countries[8,87,88]. Furthermore, the emergence of *S. flexneri* 1d, X variant (Xv) and 4s has been reported in China[89–91]. Many of these novel serotypes harbour more than one O antigen-modifying operon, resulting in additional modifications in the already highly modified tetrasaccharide. For example, the introduction of *gtr1C* into *S. flexneri* 1a leads to the addition of a glucosyl group on the glucosyl-linked *N*-acetylglucosamine, effectively converting this serotype into the novel serotype 1c[92]. Unpredictably, the *gtr1C* cluster shares similarities with genes from *Citrobacter koseri* rather than with previously characterized orthologues in other *S. flexneri* serotypes. This suggests that *S. flexneri* can sample from a large pool of O antigen-modifying genes. Plasmid-mediated serotype conversion has also been reported in *S. flexneri* Xv, 4s and Yv. The plasmid-borne O antigen phosphoethanolamine transferase (*opt*) gene was found to be essential for the transfer of phosphoethanolamine (PEtN) to the second rhamnose (RhaII) and RhaIII of the O antigen in *S. flexneri* Xv and *S. flexneri* Yv, respectively[93,94].

© 2016 Macmillan Publishers Limited. All rights reserved.

Molecular typing of *S. flexneri* has, to date, largely relied on pulsed-field gel electrophoresis (PFGE) and/or multilocus sequence typing (MLST), using the sequences of the seven housekeeping genes: *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA*[95,96]. MLST of more than 100 Asian *S. flexneri* isolates revealed that serotypes 1–5, X and Y belong to a discrete clonal complex (ST245 of the ST245 complex), whereas serotype 6 forms a distinct clonal complex (ST145 of the ST243 complex)[97,98]. Although the resolution of MLST for *S. flexneri* is limited because of an inadequate number of differentiating mutations in the selected housekeeping genes, especially for investigating local clonal expansion or fine-scaled phylogenetic relationships, this method has provided insights into the genetic relationship between major *S. flexneri* serotypes. For example, studies examining the spread of the epidemic *S. flexneri* clone ST91 in China have low resolution, but have aided the tracking of this pathogen across the region[90]. *S. flexneri* clone ST91, which was typed using another *E. coli* genotyping scheme[90], is actually typed ST245 using the *Shigella* spp. MLST approach described above[95]. The alternative *E. coli* typing scheme relies on 15 housekeeping genes — *aspC*, *clpX*, *fadD*, *icdA*, *lysP*, *mdh*, *uidA*, *arcA*, *aroE*, *cyaA*, *dnaG*, *grpE*, *mtlD*, *mutS* and *rpoS* — and provides better resolution for MLST, especially for clonal populations, such as the *S. flexneri* ST245 complex. To obtain even higher resolution, this expanded MLST scheme was combined with PFGE to investigate the expansion of *S. flexneri* clone ST91. Somewhat atypically for members of the genus *Shigella*, *S. flexneri* clone ST91 underwent at least 57 independent serotype switching events during its clonal expansion in China[90], illustrating the potential problem with using serotyping as a proxy for genetic relatedness. A major serotype conversion in the *S. flexneri* ST245 complex led to the rise of a novel variant, *S. flexneri* Xv, which then rapidly spread and became one of the most prevalent serotypes in China since 2000 (REF. 90). The spread of *S. flexneri* Xv is concerning, as this serotype is resistant to several antimicrobials (see below).

Extensive serotype switching and the success of specific clones highlight the need for higher-resolution tracking and monitoring of *S. flexneri*. Whole-genome sequencing provides such higher-resolution data; for example, this method showed that *S. flexneri* ST91 serotype Xv had acquired a plasmid carrying *opt*, leading to O antigen modification, on three independent occasions[94]. Before the *opt*-harbouring plasmid was introduced, clone ST91 had already carried antimicrobial-resistance genes, including the SRL locus (a multidrug-resistance (MDR) genomic island harbouring resistance genes against tetracycline (*tetACDR*), streptomycin (*aadA2*), ampicillin (*oxa1*) and chloramphenicol (*cat*)), Tn*7* (an MDR transposon carrying resistance genes against trimethoprim (*dfrA1*), streptothricin (*sat1*) and streptomycin (*aadA1*)) and two mutations in *gyrA* facilitating resistance against nalidixic acid. The rapid expansion of the ST91 clone in different geographical locations can be explained by O antigen switching and the evasion of pre-existing immunity in host populations, and by the ineffectiveness of antimicrobials owing to the MDR background, which promotes prolonged faecal shedding and sustained circulation[74].

In addition to the substantial species shift observed in developing countries, *S. flexneri* epidemiology has also changed in certain populations in developed countries. The isolation rate of *S. flexneri* 3a has increased steadily in men who have sex with men (MSM) communities in Canada, England and Wales[99,100]. This increased isolation rate is not attributable to an introduction (or introductions) from the low-income countries, suggesting that the ecology of this particular variant may now be better adapted to transmission within MSM populations[99]. A recent genomic analysis of a global collection of this serotype indicated the emergence of an *S. flexneri* 3a lineage attributed to infections in MSM populations in higher-income countries[101]. This lineage has spread globally since its emergence in 1998, and as is common for current populations of *Shigella* spp., has acquired resistance to multiple antimicrobials, most notably azithromycin, a frequent antimicrobial treatment for sexually transmitted diseases, including gonorrhoea, syphilis and chlamydia. This change in antimicrobial susceptibility, seen as the response to selective pressure exerted by azithromycin treatment for comorbid infections, has contributed to the dominance of this organism in MSM populations[101].

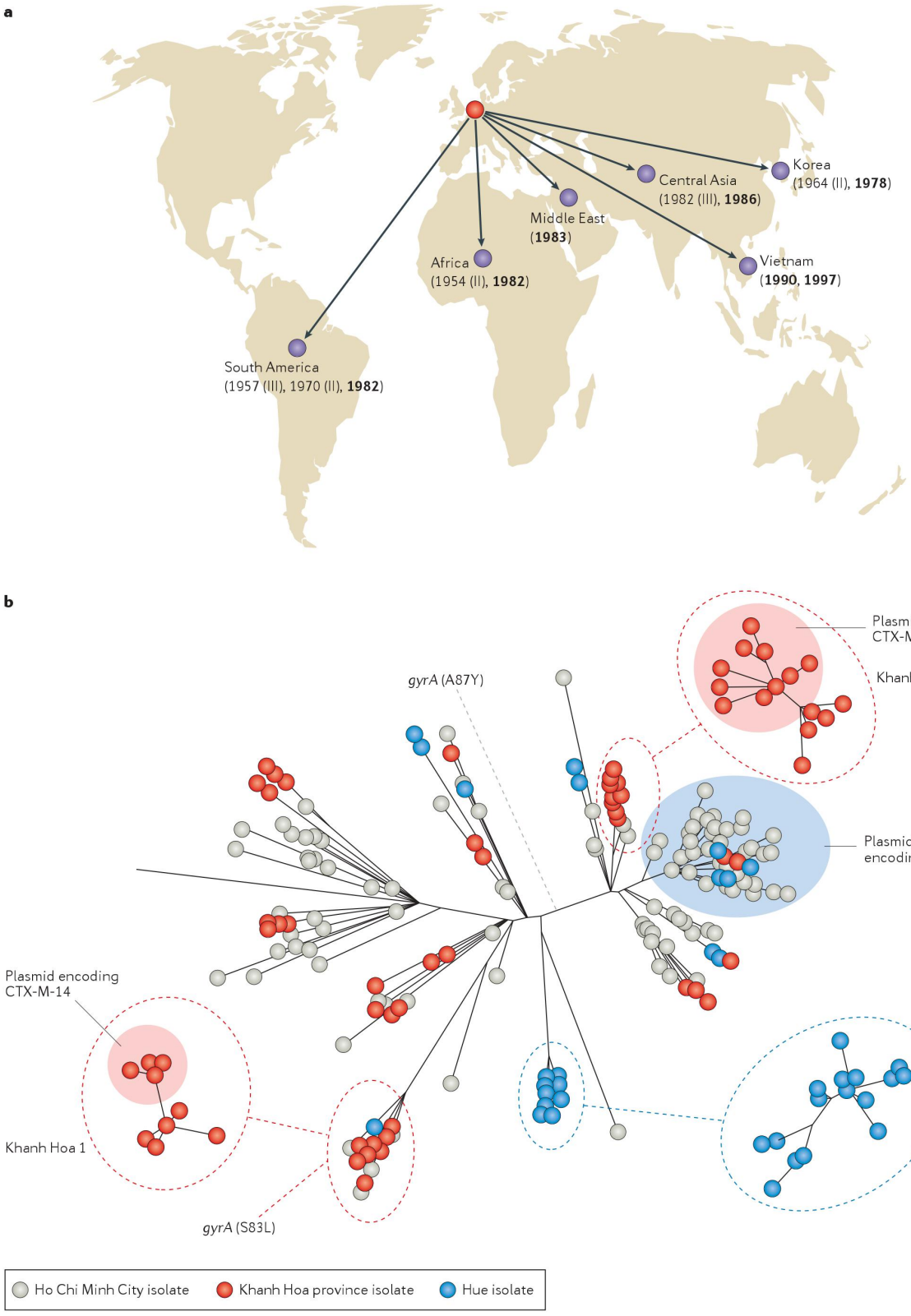
Studying the evolution and epidemiology of *S. flexneri* has proved complicated owing to serotype diversity, until a recent study of 351 whole-genome sequences from different serotypes of this species[24]. This study concluded that *S. flexneri*, with the exclusion of the diverging serotype 6, consists of seven phylogenetic groups (FIG. 3). Notably, these phylogenetic groups are inconsistent with serotype groupings and have arisen on several occasions between the 1300s and the 1800s[24]. The presence of numerous serotypes in all phylogenetic groups suggests that serotype switching is common, consistent with previous research[90].

This study also revealed substantial variability in the composition of *S. flexneri* virulence factors (for example, the genomic island SHI-1, and genes encoding iron uptake systems, such as the enterobactin genes and the ferric dicitrate transport (*fec*) locus) and antimicrobial-resistance genes (for example, the SRL island). SHI-1, SRL and enterobactin genes exclusively co-occur in phylogenetic group 3 (PG3), which is composed predominantly of *S. flexneri* serotype 2a, and this may account for the enhanced virulence and international dominance of this serotype[24]. The accumulation of antimicrobial-resistance genes in *S. flexneri* over the past three decades is considered to be essential for maintaining successful lineages. However, unlike for *S. sonnei*, this has neither led to the displacement of pre-existing antimicrobial-susceptible lineages nor resulted in substantial international transmission, with the exception of the global spread of the MSM-associated serotype *S. flexneri* 3a[22,101]. This finding supports the concept of longer-term colonization, in which diverse populations of both antimicrobial-resistant and antimicrobial-susceptible lineages co-circulate in endemic locations. These data also imply that *S. flexneri* is persisting in the environment, where selection for antimicrobial resistance may be less favourable.

Pulsed-field gel electrophoresis (PFGE). A molecular typing technique based on the migration pattern of DNA fragments of variable lengths, generated by restriction enzyme treatment, in an electrical field.

Multilocus sequence typing (MLST). A DNA sequence-based molecular typing scheme in which each isolate is distinguished by a combination of unique alleles of housekeeping genes (by comparing their genetic variations).

© 2016 Macmillan Publishers Limited. All rights reserved.



© 2016 Macmillan Publishers Limited. All rights reserved.

Homoplasies
Phenotypic or genotypic characteristics that are shared by a set of organisms but not inherited from a common ancestor.

Shigella dysenteriae *and* Shigella boydii. As *S. dysenteriae* and *S. boydii* account for <10% of the cases of shigellosis, research into these organisms is less of a priority for global health research[19]. Furthermore, research is complicated by the sheer diversity of serotypes in these species (15 for *S. dysenteriae*, and 20 for *S. boydii*) and by a lack of large, well-characterized, geographically diverse collections of isolates.

One of the best studied *S. dysenteriae* serotypes is *S. dysenteriae* serotype 1, which induces a more severe disease phenotype than other *Shigella* spp. and serotypes. The hypervirulence of *S. dysenteriae* 1 can be explained by the release of Stx, which inhibits protein synthesis[102], and could also be partly attributed to the presence of the *shu* cluster, which is upregulated in response to the host body temperature and uses haem as an iron source, leading to better adaptation in the human host[103,104]. Notably, EHEC O157:H7 also carries the *stx*-encoding prophage and the *shu* cluster and can also cause severe complications, demonstrating that *S. dysenteriae* 1 and EHEC have inherited and maintained these virulence factors from a common ancestor[103]. A detailed study of the proteomic profile of *S. dysenteriae* 1 revealed several proteins that are expressed preferentially in the host environment, including the Mxi–Spa T3SS, and proteins that are involved in anaerobic energy metabolism, acid resistance, modulation of the outer membrane and modification of peptidoglycan structure[105]. The last reported dysentery outbreak caused by *S. dysenteriae* 1 occurred in Sierra Leone in 1999 (REF. 106); since then, the prevalence of disease caused by this serotype has become negligible[7]. The most recent pandemic clone of *S. dysenteriae* 1 emerged from a common ancestor at the beginning of the twentieth century, which is much more recent than the common ancestor of the major current clones of both *S. sonnei* and *S. flexneri*[23]. Two lineages of *S. dysenteriae* 1 rapidly disseminated intercontinentally, facilitated by poor sanitation and excessive human migration during the two world wars[23]. Furthermore, these lineages independently acquired antimicrobial-resistance genes, seemingly through selection during outbreaks. However, directional selection in the chromosome is unlikely to occur, as inactivating mutations equally affected all metabolic functions[23]. Such unbiased mutations and a generally high mutation rate suggest that *S. dysenteriae* 1 could be maintained and transmitted through long-term human carriage, similarly to *S.* Typhi[107]. This theory may explain the infrequent isolation rate of *S. dysenteriae* 1 but its ability to cause devastating outbreaks in vulnerable populations.

S. boydii was first isolated in the Indian subcontinent and seems to be restricted to this region, as it is rarely isolated elsewhere[108]. However, a new serotype, *S. boydii* serotype 20, was discovered in travellers to Central America, which demonstrates that the epidemiology of *S. boydii* is more complicated than previously described or predicted[109,110]. In developing countries, *S. boydii* serotype 2 is the most prevalent and clinically relevant serotype, with an isolation rate of ~50% of all *S. boydii* isolates[111–113]. Other *S. boydii* serotypes are rare, but several O antigen clusters from *S. boydii* have been transferred to different members of the genus *Escherichia*; for example, *S. boydii* O antigens 10 and 15 can be found in EHEC and *Escherichia fergusonii*, respectively[114,115].

Linking genomics and pathogenesis. As members of the genus *Shigella* do not form a single monophyletic group, distinct *Shigella* spp. can differ in both physiology and pathogenesis. *Shigella* pathogenesis mainly relies on the Mxi–Spa T3SS and its effector proteins, so the subtle phenotypic variation seen in host–pathogen interactions could be caused by the gain and/or loss of other genetic material. Alternatively, convergent evolution has enabled several *Shigella* spp. to adopt an intracellular lifestyle, exemplified by the independent loss of flagella, fimbriae, and metabolic pathways, such as LDC, carbon utilization and transporters (of carbohydrates, amino acids and amines)[20,57,58].

Information related to pathogenic differences between and in the various *Shigella* spp. is scarce, because most experimental studies have used *S. flexneri*. The other species are used less frequently for experiments owing to the instability of their virulence plasmid (*S. sonnei*), their unavailability or simply the fact that they are less of a global health priority (*S. dysenteriae* and *S. boydii*). Nevertheless, recent findings have shed more light on variation between the different species. For example, it has been shown that bacteriophage-borne glycosylation of the O antigen in *S. flexneri* 5 optimizes its length, enhancing the exposure of the T3SS apparatus without making it more of a target for host antibodies[78]. There is a fine balance between virulence and immune protection: in *S. flexneri* serotype 2a, the plasmid pHS-2 carries a gene that results in very long O antigen chains which mask the cell from serum killing, whereas the chromosomally determined chains are short and unmask the T3SS structure to enhance functionality[116,117]. *S. sonnei* uses a different mechanism: this species expresses a group 4 capsule composed of pINV-borne O antigen sugars[118]. Removal of the capsule

◀ Figure 2 | **The intercontinental and regional dissemination of *Shigella sonnei*.** **a** | A global map showing the spread of *Shigella sonnei* out of Europe, using data from REF. 22. *S. sonnei* diverged into three main lineages (I, II and III) that have been circulating in Europe since the early nineteenth century (red). Years represent the estimated dates of introduction of these strains from Europe into new human populations. The most successful of these global lineages is lineage III, which harbours a combination of antimicrobial-resistance genes (bold dates indicate the introduction of the Global III clade). **b** | An unrooted phylogenetic tree showing the relationship between sequenced *S. sonnei* strains isolated in three different locations across Vietnam: Ho Chi Minh City in the south, Khanh Hoa province on the south-central coast and Hue in the central region; based on data from REF. 75. The tree shows that strains from Ho Chi Minh City are frequently transferred to other Vietnamese cities and rarely form new populations. However, as highlighted by two clonal expansions in Khanh Hoa (Khanh Hoa 1 and Khanh Hoa 2) and one in Hue, pioneering *S. sonnei* strains can form new location-specific subpopulations. The ongoing selection of these organisms seems to be driven by antimicrobials, as there is evidence of homoplasies by the acquisition and maintenance of differing DNA gyrase subunit A (*gyrA*) mutations and of differing plasmids encoding extended-spectrum β-lactamases (ESBLs), which confer resistance to fluoroquinolones and third-generation cephalosporins, respectively. Strains harbouring ESBL-encoding plasmids are highlighted by background shading (blue for the incompatibility group I1 (IncI1) plasmid pKHSB1 encoding CTX-M-15, and red for the IncA/C plasmid encoding CTX-M-14). Part **a** is reproduced from REF. 169, Nature Publishing Group.

© 2016 Macmillan Publishers Limited. All rights reserved.

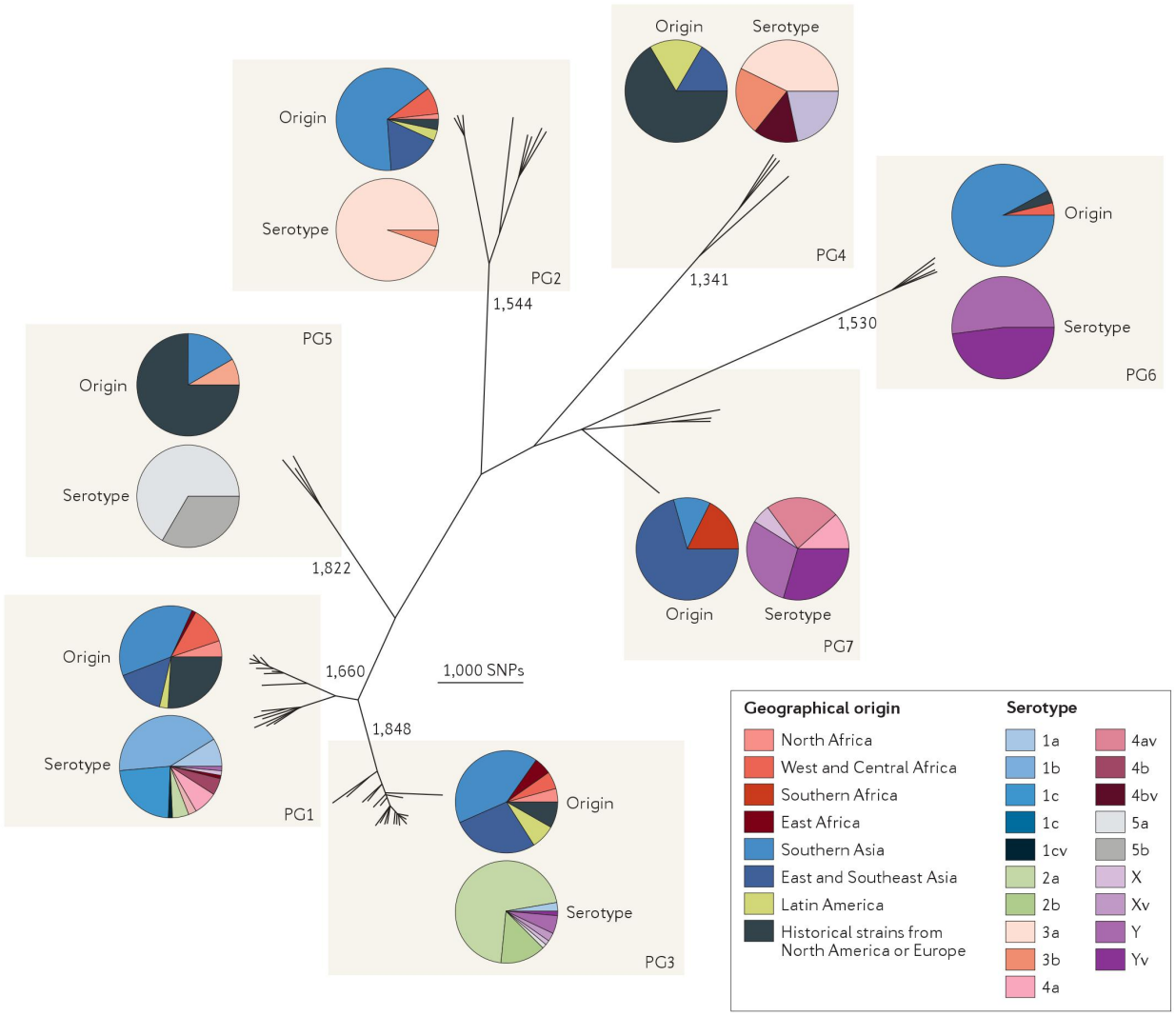


Figure 3 | **The phylogenetic structure and global distribution of the 19 *Shigella flexneri* serotypes.** The figure shows a maximum likelihood phylogeny of *Shigella flexneri* serotypes, created from genome sequences of a representative global collection of 351 isolates of S. *flexneri*, spanning serotypes F1–F5, FX, FXv, FY and FYv. The isolates were collected from the main foci of endemic disease today (Africa, Asia, and South and Central America), and historical isolates from reference collections dating back to 1914 were also used. Phylogenetic groups (PGs) determined by Bayesian analysis of population structure clustering are boxed, and the geographical and serotypic composition of isolates in each PG are inlaid as pie charts. This figure is reproduced from REF. 24.

increases invasiveness and inflammation, but decreases the capacity to spread from cell to cell and increases susceptibility to immune killing, thus showing that the capsule is crucial for the balance between virulence and immune evasion. The *g4c* operon, which encodes this capsule, is intact in *S. sonnei* but is lacking in *S. flexneri* serotype 2a owing to a frame-shift deletion[118]. This variance may explain, in part, the differential virulence and immunogenicity of *Shigella* spp. Differences also exist in the use of adhesins for attachment to host cells. In *S. flexneri*, the T3SS-dependent protein IcsA (also known as VirG) mediates both adhesion and actin-based motility, facilitating invasion into the host cell[119]. In *S. sonnei*, an additional multivalent adhesion molecule (MAM), SSO1327, has been shown to function as a non-redundant adhesin to IcsA[120]. Deletion of either of these proteins in *S. sonnei* reduces attachment and invasion *in vivo*. The gene encoding SSO1327 is intact in isolates of *S. sonnei*, *S. dysenteriae* and *S. boydii*, but is a pseudogene in *S. flexneri*[120]. This difference in adhesin composition may explain the differential interaction of *S. flexneri* and *S. sonnei* with the host; for example, bile salts stimulate the attachment of *S. flexneri* but impede the attachment of *S. sonnei*[119,120].

© 2016 Macmillan Publishers Limited. All rights reserved.

Conclusions and outlook

The evolutionary history of the bacterial genus *Shigella* is shaped by three key processes. First, *Shigella* spp. have arisen from different ancestral *E. coli* strains on several independent occasions. Second, the acquisition of plasmids that encode virulence genes into numerous ancestral *Shigella* strains were 'foothold moments' in their evolution; similar observations have been made for other enteric human pathogens, such as *Y. pestis* and, more recently, *Yersinia enterocolitica*[121]. The acquisition and adaptation of these plasmids has shaped all existing *Shigella* spp. Third, convergent evolution, by the independent acquisition of mobile elements and loss of gene function, has further transformed these organisms to become restricted to humans and exquisitely customized to interact with the human intestinal mucosa.

The shift in dominance from *S. flexneri* to *S. sonnei* in economically transitioning nations should prompt more in-depth studies of the evolution and epidemiology of these two species. Although whole-genome analyses of *S. sonnei* and *S. flexneri* provided insights into their evolution and spread, comparatively little is currently understood about *S. dysenteriae* and *S. boydii*. As genome sequencing becomes more accessible and affordable, it will be essential to apply this tool to investigate the evolution of other *Shigella* spp. locally and globally. Greater insights into the epidemiology of these species should aid their control in disease-burdened regions as well as facilitate vaccine development and distribution. Conserved proteins across all *Shigella* spp., such as the T3SS proteins IpaB and IpaD, have been identified as promising candidates for a serotype-independent pan-*Shigella* vaccine. Preclinical testing in mice indicates that IpaB and IpaD are safe and provide substantial protection against challenges with *S. flexneri* and *S. sonnei*[122–124]. However, the utility of these antigens needs to be further validated in human studies. Owing to the multiple serotypes of *S. flexneri*, their complex evolutionary history and the extent of horizontal gene transfer, studying this species is more challenging. Further, *S. boydii* and *S. dysenteriae* research has been neglected owing to their lower disease burdens. *S. dysenteriae* serotype 1, in particular, warrants more attention because it can cause severe disease and has the potential to cause major epidemics. Future laboratory research should be integrated with genomics to address the survival, transmission and evolution of *Shigella* spp., focusing on how their lifestyle in the environment can affect disease epidemiology and global public health.

1. World Health Organization. *World Health Statistics 2014*. (WHO, 2014).
2. Liu, L. *et al.* Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet* **379**, 2151–2161 (2012).
3. Fischer Walker, C. L., Perin, J., Aryee, M. J., Boschi-Pinto, C. & Black, R. E. Diarrhea incidence in low- and middle-income countries in 1990 and 2010: a systematic review. *BMC Publ. Health* **12**, 220 (2012).
4. Kotloff, K. L. *et al.* The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiologic and clinical methods of the case/control study. *Clin. Infect. Dis.* **55** (Suppl. 4), S232–S245 (2012).
5. Kotloff, K. L. *et al.* Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case–control study. *Lancet* **382**, 209–222 (2013).
6. Kotloff, K. L. *et al.* Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. *Bull. World Health Organ.* **77**, 651–666 (1999).
7. Bardhan, P., Faruque, A. S., Naheed, A. & Sack, D. A. Decrease in shigellosis-related deaths without *Shigella* spp.-specific interventions, Asia. *Emerg. Infect. Dis.* **16**, 1718–1723 (2010).
8. von Seidlein, L. *et al.* A multicentre study of *Shigella* diarrhoea in six Asian countries: disease burden, clinical manifestations, and microbiology. *PLoS Med.* **3**, e353 (2006).
9. Nataro, J. P. & Kaper, J. B. Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.* **11**, 142–201 (1998).
10. Croxen, M. A. & Finlay, B. B. Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat. Rev. Microbiol.* **8**, 26–38 (2010).
11. Croxen, M. A. *et al.* Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clin. Microbiol. Rev.* **26**, 822–880 (2013).
12. Kothary, M. & Babu, U. Infective dose of foodborne pathogens in volunteers: a review. *J. Food Saf.* **21**, 49–68 (2001).
13. Shiga, K. Ueber den erreger der dysenterie in Japan. *Zentralbl. Bakteriol. Mikrobiol.* **23**, 599–600 (1898).
14. Trofa, A. F., Ueno-olsen, H., Oiwa, R. & Yoshikawa, M. Dr. Kiyoshi Shiga: discoverer of the dysentery *Bacillus*. *Clin. Infect. Dis.* **29**, 1303–1306 (1999).
15. Flexner, S. On the etiology of tropical dysentery. *Philadelphia Med. J.* **6**, 414–421 (1900).
16. Barceloux, D. G. *Shigella species (Shiga enterotoxins) in Medical Toxicology of Natural Substances: Foods, Fungi, Medicinal Herbs, Plants, and Venomous Animals*. Ch. 20, 150–155. (Wiley & Sons, 2008).
17. Dodd, C. E. & Jones, D. A numerical taxonomic study of the genus *Shigella*. *J. Gen. Microbiol.* **128**, 1933–1957 (1982).
18. Falkow, S. Activity of lysine decarboxylase as an aid in the identification of *Salmonellae* and *Shigellae*. *Am. J. Clin. Pathol.* **29**, 598–600 (1958).
19. Gu, B. *et al.* Comparison of the prevalence and changing resistance to nalidixic acid and ciprofloxacin of *Shigella* between Europe–America and Asia–Africa from 1998 to 2009. *Int. J. Antimicrob. Agents* **40**, 9–17 (2012).
20. Yang, F. *et al.* Genome dynamics and diversity of *Shigella* species, the etiologic agents of bacillary dysentery. *Nucleic Acids Res.* **33**, 6445–6458 (2005).
A seminal paper aimed at genomic comparison of four *Shigella* spp.: *S. flexneri*, *S. sonnei*, *S. dysenteriae* and *S. boydii*. The analysis provides evidence for convergent evolution among *Shigella* spp., by gene gain and gene loss.
21. Buchrieser, C. *et al.* The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. *Mol. Microbiol.* **38**, 760–771 (2000).
22. Holt, K. E. *et al. Shigella sonnei* genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. *Nat. Genet.* **44**, 1056–1059 (2012).
This key investigation reconstructed the phylogenetic structure and geographical spread of *S. sonnei* on a global scale, emphasizing that the successful introduction of this species from Europe into other distant regions is facilitated by antimicrobial resistance.
23. Rohmer, L. *et al.* Genomic analysis of the emergence of 20th century epidemic dysentery. *BMC Genomics* **15**, 355 (2014).
The first work to use phylogenomics to understand the evolutionary history of the pandemic *S. dysenteriae* 1, highlighting the fact that its recent emergence is coupled with independent acquisitions of antimicrobial-resistance genes.
24. Connor, T. R. *et al.* Species-wide whole genome sequencing reveals historical global spread and recent local persistence in *Shigella flexneri*. *eLife* **4**, e07335 (2015).
The original whole-genome study on the diversity of *S. flexneri* globally, showing that serotype switching is common among phylogenetic groups. The study suggests that the long-term success of this species is linked to its local persistence.
25. Lan, R., Lumb, B., Ryan, D. & Reeves, P. R. Molecular evolution of large virulence plasmid in *Shigella* clones and enteroinvasive *Escherichia coli*. *Infect. Immun.* **69**, 6303–6309 (2001).
26. Pupo, G. M., Lan, R. & Reeves, P. R. Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc. Natl Acad. Sci. USA* **97**, 10567–10572 (2000).
The first article to demonstrate the multiple origins of *Shigella* spp. from *E. coli*, on the basis of a phylogenetic analysis of eight housekeeping genes.
27. Yang, J. *et al.* Revisiting the molecular evolutionary history of *Shigella* spp. *J. Mol. Evol.* **64**, 71–79 (2007).
28. Hyma, K. E. *et al.* Evolutionary genetics of a new pathogenic *Escherichia* species: *Escherichia albertii* and related *Shigella boydii* strains. *J. Bacteriol.* **187**, 619–628 (2005).
29. Shepherd, J. G., Wang, L. & Reeves, P. R. Comparison of O-antigen gene clusters of *Escherichia coli* (*Shigella*) *sonnei* and *Plesiomonas shigelloides* O17: *sonnei* gained its current plasmid-borne O-antigen genes from *P. shigelloides* in a recent event. *Infect. Immun.* **68**, 6056–6061 (2000).
30. Sims, G. E. & Kim, S. H. Whole-genome phylogeny of *Escherichia coli*/*Shigella* group by feature frequency profiles (FFPs). *Proc. Natl Acad. Sci. USA* **108**, 8329–8334 (2011).
31. Touchon, M. *et al.* Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* **5**, e1000344 (2009).
32. Kaas, R. S., Friis, C., Ussery, D. W. & Aarestrup, F. M. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. *BMC Genomics* **13**, 577 (2012).
33. Zhang, Y. & Lin, K. A phylogenomic analysis of *Escherichia coli* / *Shigella* group: implications of genomic features associated with pathogenicity and ecological adaptation. *BMC Evol. Biol.* **12**, 174 (2012).

© 2016 Macmillan Publishers Limited. All rights reserved.

34. Sahl, J. W. *et al.* Defining the phylogenomics of *Shigella* species: a pathway to diagnostics. *J. Clin. Microbiol.* **53**, 951–960 (2015).
This paper utilizes the most complete collection of *Shigella* and *E. coli* isolates for phylogenomic analysis, the results of which both consolidate the evidence for there being multiple origins of *Shigella* spp. from *E. coli*, and point to reliable genetic markers for differentiating major *Shigella* clades.
35. van den Beld, M. J. & Reubsaet, F. A. Differentiation between *Shigella*, enteroinvasive *Escherichia coli* (EIEC) and noninvasive *Escherichia coli*. *Eur. J. Clin. Microbiol. Infect. Dis.* **31**, 899–904 (2012).
36. Lan, R. *et al.* Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and *Shigella* spp. *Infect. Immun.* **72**, 5080–5088 (2004).
37. Schroeder, G. N. & Hilbi, H. Molecular pathogenesis of *Shigella* spp.: controlling host cell signaling, invasion, and death by type III secretion. *Clin. Microbiol. Rev.* **21**, 134–156 (2008).
This extensive review details the molecular pathogenesis of *Shigella* spp.
38. Venkatesan, M. M. *et al.* Complete DNA sequence and analysis of the large virulence plasmid of *Shigella flexneri* complete DNA sequence and analysis of the large virulence plasmid of *Shigella flexneri*. *Infect. Immun.* **69**, 3271–3285 (2001).
Together with reference 21, this report describes the first sequence analysis of the *Shigella* virulence plasmid, highlighting the rich repertoire of virulence factors encoded by this plasmid.
39. Jiang, Y. *et al.* The complete sequence and analysis of the large virulence plasmid pSS of *Shigella sonnei*. *Plasmid* **54**, 149–159 (2005).
40. Makino, S., Sasakawa, C. & Yoshikawa, M. Genetic relatedness of the basic replicon of the virulence plasmid in shigellae and enteroinvasive *Escherichia coli*. *Microb. Pathog.* **5**, 267–274 (1988).
41. Lan, R. & Reeves, P. R. *Escherichia coli* in disguise: molecular origins of *Shigella*. *Microbes Infect.* **4**, 1125–1132 (2002).
42. Al-Hasani, K. *et al.* The *sigA* gene which is borne on the *she* pathogenicity island of *Shigella flexneri* 2a encodes an exported cytopathic protease involved in intestinal fluid accumulation. *Infect. Immun.* **68**, 2457–2463 (2000).
43. Ingersoll, M., Groisman, E. A. & Zychlinsky, A. Pathogenicity islands of *Shigella*. *Curr. Top. Microbiol. Immunol.* **264**, 49–65 (2002).
44. Luck, S. N., Turner, S. A., Rajakumar, K., Sakellaris, H. & Adler, B. Ferric dicitrate transport system (Fec) of *Shigella flexneri* 2a YSH6000 is encoded on a novel pathogenicity island carrying multiple antibiotic resistance genes. *Infect. Immun.* **69**, 6012–6021 (2001).
45. Vokes, S. A., Reeves, S. A., Torres, A. G. & Payne, S. M. The aerobactin iron transport system genes in *Shigella flexneri* are present within a pathogenicity island. *Mol. Microbiol.* **33**, 63–73 (1999).
46. Fisher, C. R. *et al.* Genetics and virulence association of the *Shigella flexneri* Sit iron transport system. *Infect. Immun.* **77**, 1992–1999 (2009).
47. Gupta, S. K. *et al.* Short report: emergence of Shiga toxin 1 genes within *Shigella dysenteriae* type 4 isolates from travelers returning from the Island of Hispañola. *Am. J. Trop. Med. Hyg.* **76**, 1163–1165 (2007).
48. Gray, M. D. *et al.* Clinical isolates of Shiga toxin 1a-producing *Shigella flexneri* with an epidemiological link to recent travel to Hispañiola. *Emerg. Infect. Dis.* **20**, 1669–1677 (2014).
49. Gray, M. D. *et al.* Prevalence of Shiga toxin-producing *Shigella* species isolated from French travellers returning from the Caribbean: an emerging pathogen with international implications. *Clin. Microbiol. Infect.* **21**, 765.e9–765.e14 (2015).
50. Gray, M. *et al.* Stx-producing *Shigella* species from patients in Haiti: an emerging pathogen with the potential for global spread. *Open Forum Infect. Dis.* **2**, ofv134 (2015).
51. Siguier, P., Gourbeyre, E. & Chandler, M. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol. Rev.* **38**, 865–891 (2014).
52. Wagner, A. & de la Chaux, N. Distant horizontal gene transfer is rare for multiple families of prokaryotic insertion sequences. *Mol. Genet. Genom.* **280**, 397–408 (2008).
53. Ochman, H. & Davalos, L. M. The nature and dynamics of bacterial genomes. *Science* **311**, 1730–1733 (2006).
54. Eiglmeier, K. *et al.* The decaying genome of *Mycobacterium leprae*. *Lepr. Rev.* **72**, 387–398 (2001).
55. Parkhill, J. *et al.* Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523–527 (2001).
56. Holt, K. E. *et al.* High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. *Nat. Genet.* **40**, 987–993 (2008).
57. Feng, Y., Chen, Z. & Liu, S.-L. Gene decay in *Shigella* as an incipient stage of host-adaptation. *PLoS ONE* **6**, e27754 (2011).
58. Bravo, V., Puhar, A., Sansonetti, P., Parsot, C. & Toro, C. S. Distinct mutations led to inactivation of type 1 fimbriae expression in *Shigella* spp. *PLoS ONE* **10**, e0121785 (2015).
59. Ramos, H. C., Rumbo, M. & Sirard, J. C. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. *Trends Microbiol.* **12**, 509–517 (2004).
60. Monk, J. M. *et al.* Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proc. Natl Acad. Sci. USA* **110**, 20338–20343 (2013).
61. Maurelli, A. T., Fernandez, R. E., Bloch, C. A., Rode, C. K. & Fasano, A. 'Black holes' and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **95**, 3943–3948 (1998).
62. Prunier, A. L. *et al. nadA* and *nadB* of *Shigella flexneri* 5a are antivirulence loci responsible for the synthesis of quinolinate, a small molecule inhibitor of *Shigella* pathogenicity. *Microbiology* **153**, 2363–2372 (2007).
63. Prosseda, G. *et al.* Shedding of genes that interfere with the pathogenic lifestyle: the *Shigella* model. *Res. Microbiol.* **163**, 399–406 (2012).
64. Nakata, N. *et al.* The absence of a surface protease, OmpT, determines the intercellular spreading ability of *Shigella*: the relationship between the *ompT* and *kcpA* loci. *Mol. Microbiol.* **9**, 459–468 (1993).
65. Zhao, G. *et al.* A novel anti-virulence gene revealed by proteomic analysis in *Shigella flexneri* 2a. *Proteome Sci.* **8**, 30 (2010).
66. Barbagallo, M. *et al.* A new piece of the *Shigella* pathogenicity puzzle: spermidine accumulation by silencing of the *speG* gene. *PLoS ONE* **6**, e27226 (2011).
67. Hershberg, R., Tang, H. & Petrov, D. A. Reduced selection leads to accelerated gene loss in *Shigella*. *Genome Biol.* **8**, R164 (2007).
68. Balbi, K. J., Rocha, E. P. C. & Feil, E. J. The temporal dynamics of slightly deleterious mutations in *Escherichia coli* and *Shigella* spp. *Mol. Biol. Evol.* **26**, 345–355 (2009).
69. Vinh, H. *et al.* A changing picture of shigellosis in southern Vietnam: shifting species dominance, antimicrobial susceptibility and clinical presentation. *BMC Infect. Dis.* **9**, 204 (2009).
70. Ud-Din, A. I. M. S. *et al.* Changing trends in the prevalence of *Shigella* species: emergence of multi-drug resistant *Shigella sonnei* biotype g in Bangladesh. *PLoS ONE* **8**, e82601 (2013).
71. Bangtrakulnonth, A. *et al. Shigella* from humans in Thailand during 1993 to 2006: spatial-time trends in species and serotype distribution. *Foodborne Pathog. Dis.* **5**, 773–784 (2008).
72. Wei, H. L., Wang, Y. W., Li, C. C., Tung, S. K. & Chiou, C. S. Epidemiology and evolution of genotype and antimicrobial resistance of an imported *Shigella sonnei* clone circulating in central Taiwan. *Diagn. Microbiol. Infect. Dis.* **58**, 469–475 (2007).
73. Nygren, B. L. *et al.* Foodborne outbreaks of shigellosis in the USA, 1998–2008. *Epidemiol. Infect.* **141**, 233–241 (2013).
74. Vinh, H. *et al.* Treatment of bacillary dysentery in Vietnamese children: two doses of ofloxacin versus 5-days nalidixic acid. *Trans. R. Soc. Trop. Med. Hyg.* **94**, 323–326 (2000).
75. Holt, K. E. *et al.* Tracking the establishment of local endemic populations of an emergent enteric pathogen. *Proc. Natl Acad. Sci. USA* **110**, 17522–17527 (2013).
The first study to detail the microevolution of *S. sonnei* in a developing region in high resolution, pointing out that antimicrobial resistance is a crucial factor for understanding the clonal expansion of this species.
76. Nguyen, N. T. K. *et al.* The sudden dominance of bla_{CTX-M} harbouring plasmids in *Shigella* spp. circulating southern Vietnam. *PLoS Negl. Trop. Dis.* **4**, e702 (2010).
77. Rashid, H. & Rahman, M. Possible transfer of plasmid mediated third generation cephalosporin resistance between *Escherichia coli* and *Shigella sonnei* in the human gut. *Infect. Genet. Evol.* **30**, 15–18 (2015).
78. West, N. P. *et al.* Optimization of virulence functions through glucosylation of *Shigella* LPS. *Science* **307**, 1313–1317 (2005).
79. Allison, G. E. & Verma, N. K. Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*. *Trends Microbiol.* **8**, 17–23 (2000).
80. Bastin, D. A., Lord, A. & Verma, N. K. Cloning and analysis of the glucosyl transferase gene encoding type I antigen in *Shigella flexneri*. *FEMS Microbiol. Lett.* **156**, 133–139 (1997).
81. Sun, Q. *et al.* Isolation and genomic characterization of SfI, a serotype-converting bacteriophage of *Shigella flexneri*. *BMC Microbiol.* **13**, 39 (2013).
82. Mavris, M., Manning, P. A. & Morona, R. Mechanism of bacteriophage SfII-mediated serotype conversion in *Shigella flexneri*. *Mol. Microbiol.* **26**, 939–950 (1997).
83. Clark, C., Beltrame, J. & Manning, P. The *oac* gene encoding a lipopolysaccharide O-antigen acetylase maps adjacent to the integrase-encoding gene on the genome of *Shigella flexneri* bacteriophage Sf6. *Gene* **107**, 43–52 (1991).
84. Jakhetia, R., Talukder, K. A. & Verma, N. K. Isolation, characterization and comparative genomics of bacteriophage SfIV: a novel serotype converting phage from *Shigella flexneri*. *BMC Genomics* **14**, 677 (2013).
85. Huan, P. T., Bastin, D. A., Whittle, B. L., Lindberg, A. A. & Verma, N. K. Molecular characterization of the genes involved in O-antigen modification, attachment, integration and excision in *Shigella flexneri* bacteriophage SfV. *Gene* **195**, 217–227 (1997).
86. Guan, S., Bastin, D. & Verma, N. Functional analysis of the O antigen glucosylation gene cluster of *Shigella flexneri* bacteriophage SfX. *Microbiology* **145**, 1263–1273 (1999).
87. Wehler, T. & Carlin, N. I. Structural and immunochemical studies of the lipopolysaccharide from a new provisional serotype of *Shigella flexneri*. *Eur. J. Biochem.* **176**, 471–476 (1988).
88. Stagg, R. M., Cam, P. D. & Verma, N. K. Identification of newly recognized serotype 1c as the most prevalent *Shigella flexneri* serotype in northern rural Vietnam. *Epidemiol. Infect.* **136**, 1134–1140 (2008).
89. Luo, X. *et al.* Emergence of a novel *Shigella flexneri* serotype 1d in China. *Diagn. Microbiol. Infect. Dis.* **74**, 316–319 (2012).
90. Ye, C. *et al.* Emergence of a new multidrug-resistant serotype X variant in an epidemic clone of *Shigella flexneri*. *J. Clin. Microbiol.* **48**, 419–426 (2010).
91. Qiu, S. *et al.* Emergence of a novel *Shigella flexneri* serotype 4s strain that evolved from a serotype X variant in China. *J. Clin. Microbiol.* **49**, 1148–1150 (2011).
92. Stagg, R. M. *et al.* A novel glucosyltransferase involved in O-antigen modification of *Shigella flexneri* serotype 1c. *J. Bacteriol.* **191**, 6612–6617 (2009).
93. Sun, Q. *et al.* Identification and characterization of a novel *Shigella flexneri* serotype Yv in China. *PLoS ONE* **8**, e70238 (2013).
94. Zhang, N. *et al.* Genomic portrait of the evolution and epidemic spread of a recently emerged multidrug-resistant *Shigella flexneri* clone in China. *J. Clin. Microbiol.* **52**, 1119–1126 (2014).
95. Lacher, D. W., Steinsland, H., Blank, T. E., Donnenberg, M. S. & Whittam, T. S. Molecular evolution of typical enteropathogenic *Escherichia coli*: clonal analysis by multilocus sequence typing and virulence gene allelic profiling. *J. Bacteriol.* **189**, 342–350 (2007).
96. Ribot, E. M. *et al.* Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157:H7, *Salmonella*, and *Shigella* for PulseNet. *Foodborne Pathog. Dis.* **3**, 59–67 (2006).
97. Wirth, T. *et al.* Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol. Microbiol.* **60**, 1136–1151 (2006).
98. Choi, S. Y. *et al.* Multilocus sequence typing analysis of *Shigella flexneri* isolates collected in Asian countries. *J. Med. Microbiol.* **56**, 1460–1466 (2007).

© 2016 Macmillan Publishers Limited. All rights reserved.

99. Borg, M. L. *et al.* Ongoing outbreak of *Shigella flexneri* serotype 3a in men who have sex with men in England and Wales, data from 2009–2011. *Euro Surveill.* **17**, 20137 (2012).
100. Ratnayake, R. Allard, R. & Pilon, P. A. Shifting dominance of *Shigella* species in men who have sex with men. *Epidemiol. Infect.* **140**, 2082–2086 (2012).
101. Baker, K. S. *et al.* Intercontinental dissemination of azithromycin-resistant shigellosis through sexual transmission: a cross-sectional study. *Lancet Infect. Dis.* **15**, 913–921 (2015).
102. Khan, W. A., Griffiths, J. K. & Bennish, M. L. Gastrointestinal and extra-intestinal manifestations of childhood shigellosis in a region where all four species of *Shigella* are endemic. *PLoS ONE* **8**, e64097 (2013).
103. Wyckoff, E. E. *et al.* Structure of the *Shigella dysenteriae* haem transport locus and its phylogenetic distribution in enteric bacteria. *Mol. Microbiol.* **28**, 1139–1152 (1998).
104. Kouse, A. B., Righetti, F., Kortmann, J., Narberhaus, F. & Murphy, E. R. RNA-mediated thermoregulation of iron-acquisition genes in *Shigella dysenteriae* and pathogenic *Escherichia coli*. *PLoS ONE* **8**, e63781 (2013).
105. Kuntumalla, S. *et al. In vivo* versus *in vitro* protein abundance analysis of *Shigella dysenteriae* type 1 reveals changes in the expression of proteins involved in virulence, stress and energy metabolism. *BMC Microbiol.* **11**, 147 (2011).
106. Guerin, P. J. *et al. Shigella dysenteriae* serotype 1 in west Africa: intervention strategy for an outbreak in Sierra Leone. *Lancet* **362**, 705–706 (2003).
107. Roumagnac, P. *et al.* Evolutionary history of *Salmonella* Typhi. *Science* **314**, 1301–1304 (2006).
108. Khan, M. U., Roy, N. C., Islam, R., Huq, I. & Stoll, B. Fourteen years of shigellosis in Dhaka: an epidemiological analysis. *Int. J. Epidemiol.* **14**, 607–613 (1985).
109. Kalluri, P. *et al.* Epidemiological features of a newly described serotype of *Shigella boydii*. *Epidemiol. Infect.* **132**, 579–583 (2004).
110. Woodward, D. L. *et al.* Identification and characterization of *Shigella boydii* 20 serovar nov., a new and emerging *Shigella* serotype. *J. Med. Microbiol.* **54**, 741–748 (2005).
111. Smith, A. M. *et al.* Analysis of a temporal cluster of *Shigella boydii* isolates in Mpumalanga, South Africa, November to December 2007. *J. Infect. Dev. Ctries* **3**, 65–70 (2009).
112. El-Gendy, A. M. *et al.* Genetic diversity and antibiotic resistance in *Shigella dysenteriae* and *Shigella boydii* strains isolated from children aged <5 years in Egypt. *Epidemiol. Infect.* **140**, 299–310 (2012).
113. Livio, S. *et al. Shigella* isolates from the global enteric multicenter study inform vaccine development. *Clin. Infect. Dis.* **59**, 933–941 (2014).
Using data from the largest prospective paediatric diarrhoea study, this paper provides an up-to-date description for the prevalence of different *Shigella* spp. in developing countries.
114. Iguchi, A., Iyoda, S., Seto, K. & Ohnishi, M. Emergence of a novel Shiga toxin-producing *Escherichia coli* O serogroup cross-reacting with *Shigella boydii* type 10. *J. Clin. Microbiol.* **49**, 3678–3680 (2011).
115. Azmuda, N. *et al.* Evidence of interspecies O antigen gene cluster transfer between *Shigella boydii* 15 and *Escherichia fergusonii*. *Apmis* **120**, 959–966 (2012).
116. Hong, M. & Payne, S. M. Effect of mutations in *Shigella flexneri* chromosomal and plasmid-encoded lipopolysaccharide genes on invasion and serum resistance. *Mol. Microbiol.* **24**, 779–791 (1997).
117. Morona, R., Daniels, C. & Van Den Bosch, L. Genetic modulation of *Shigella flexneri* 2a lipopolysaccharide O antigen modal chain length reveals that it has been optimized for virulence. *Microbiology* **149**, 925–939 (2003).
118. Caboni, M. *et al.* An O antigen capsule modulates bacterial pathogenesis in *Shigella sonnei*. *PLoS Pathog.* **11**, e1004749 (2015).
119. Brotcke Zumsteg, A., Goosmann, C., Brinkmann, V., Morona, R. & Zychlinsky, A. IcsA is a *Shigella flexneri* adhesin regulated by the type III secretion system and required for pathogenesis. *Cell Host Microbe* **15**, 435–445 (2014).
120. Mahmoud, R. Y. *et al.* The multivalent adhesion molecule SSO1327 plays a key role in *Shigella sonnei* pathogenesis. *Mol. Microbiol.* **99**, 658–673 (2016).
121. Reuter, S. *et al.* Parallel independent evolution of pathogenicity within the genus *Yersinia*. *Proc. Natl Acad. Sci. USA* **111**, 6768–6773 (2014).
122. Martinez-Becerra, F. J. *et al.* Broadly protective *Shigella* vaccine based on type III secretion apparatus proteins. *Infect. Immun.* **80**, 1222–1231 (2012).
123. Heine, S. J. *et al.* Intradermal delivery of *Shigella* IpaB and IpaD type III secretion proteins: kinetics of cell recruitment and antigen uptake, mucosal and systemic immunity, and protection across serotypes. *J. Immunol.* **192**, 1630–1640 (2014).
124. Heine, S. J. *et al. Shigella* IpaB and IpaD displayed on *L. lactis* bacterium-like particles induce protective immunity in adult and infant mice. *Immunol. Cell Biol.* **93**, 641–652 (2015).
125. Carayol, N. & Tran Van Nhieu, G. Tips and tricks about *Shigella* invasion of epithelial cells. *Curr. Opin. Microbiol.* **16**, 32–37 (2013).
126. Sansonetti, P. J. *et al.* Infection of rabbit Peyer's patches by *Shigella flexneri*: effect of adhesive or invasive bacterial phenotypes on follicle-associated epithelium. *Infect. Immun.* **64**, 2752–2764 (1996).
127. Zychlinsky, A., Prevost, M. & Sansonetti, P. *Shigella flexneri* induces apoptosis in infected macrophages. *Nature* **358**, 167–169 (1992).
128. Hilbi, H. *et al. Shigella*-induced apoptosis is dependent on caspase-1 which binds to IpaB. *J. Biol. Chem.* **273**, 32895–32900 (1998).
129. Lafont, F., Tran Van Nhieu, G., Hanada, K., Sansonetti, P. & van der Goot, F. G. Initial steps of *Shigella* infection depend on the cholesterol/sphingolipid raft-mediated CD44–IpaB interaction. *EMBO J.* **21**, 4449–4457 (2002).
130. Skoudy, A. *et al.* CD44 binds to the *Shigella* IpaB protein and participates in bacterial invasion of epithelial cells. *Cell. Microbiol.* **2**, 19–33 (2000).
131. Watarai, M., Funato, S. & Sasakawa, C. Interaction of Ipa proteins of *Shigella flexneri* with $\alpha_5\beta_1$ integrin promotes entry of the bacteria into mammalian cells. *J. Exp. Med.* **183**, 991–999 (1996).
132. Mounier, J. *et al.* The IpaC carboxyterminal effector domain mediates Src-dependent actin polymerization during *Shigella* invasion of epithelial cells. *PLoS Pathog.* **5**, e1000271 (2009).
133. Handa, Y. *et al. Shigella* IpgB1 promotes bacterial entry through the ELMO–Dock180 machinery. *Nat. Cell Biol.* **9**, 121–128 (2007).
134. Niebuhr, K. *et al.* Conversion of PtdIns(4,5)P_2 into PtdIns(5)P by the *S. flexneri* effector IpgD reorganizes host cell morphology. *EMBO J.* **21**, 5069–5078 (2002).
135. Park, H., Valencia-Gallardo, C., Sharff, A., Tran Van Nhieu, G. & Izard, T. Novel vinculin binding site of the IpaA invasin of *Shigella*. *J. Biol. Chem.* **286**, 23214–23221 (2011).
136. Yoshida, S. *et al. Shigella* deliver an effector protein to trigger host microtubule destabilization, which promotes Rac1 activity and efficient bacterial internalization. *EMBO J.* **21**, 2923–2935 (2002).
137. Romero, S. *et al.* ATP-mediated Erk1/2 activation stimulates bacterial capture by filopodia, which precedes *Shigella* invasion of epithelial cells. *Cell Host Microbe* **9**, 508–519 (2011).
138. Fernandez-Prada, C. M. *et al. Shigella flexneri* $IpaH_{7.8}$ facilitates escape of virulent bacteria from the endocytic vacuoles of mouse and human macrophages. *Infect. Immun.* **68**, 3608–3619 (2000).
139. Harrington, A., Darboe, N. & Kenjale, R. Characterization of the interaction of single tryptophan containing mutants of IpaC from *Shigella flexneri* with phospholipid membranes. *Biochemistry* **45**, 626–636 (2006).
140. Egile, C., Loisel, T. & Laurent, V. Activation of the CDC42 effector N-WASP by the *Shigella flexneri* IcsA protein promotes actin nucleation by Arp2/3 complex and bacterial actin-based motility. *J. Cell Biol.* **146**, 1319–1332 (1999).
141. Iwai, H. *et al.* A bacterial effector targets Mad2L2, an APC inhibitor, to modulate host cell cycling. *Cell* **130**, 611–623 (2007).
142. Pendaries, C. *et al.* PtdIns5P activates the host cell PI3-kinase/Akt pathway during *Shigella flexneri* infection. *EMBO J.* **25**, 1024–1034 (2006).
143. Kim, M. *et al.* Bacteria hijack integrin-linked kinase to stabilize focal adhesions and block cell detachment. *Nature* **459**, 578–582 (2009).
144. Mounier, J. *et al. Shigella* effector IpaB-induced cholesterol relocation disrupts the Golgi complex and recycling network to inhibit host cell secretion. *Cell Host Microbe* **12**, 381–389 (2012).
145. Dong, N. *et al.* Structurally distinct bacterial TBC-like GAPs link Arf GTPase to Rab1 inactivation to counteract host defenses. *Cell* **150**, 1029–1041 (2012).
146. Fasano, A., Noriega, F. R., Liao, F. M., Wang, W. & Levine, M. M. Effect of *Shigella* enterotoxin 1 (ShET1) on rabbit intestine *in vitro* and *in vivo*. *Gut* **40**, 505–511 (1997).
147. Nataro, J. P. *et al.* Identification and cloning of a novel plasmid-encoded enterotoxin of enteroinvasive *Escherichia coli* and *Shigella* strains. *Infect. Immun.* **63**, 4721–4728 (1995).
148. Faherty, C. S. *et al.* Chromosomal and plasmid-encoded factors of *Shigella flexneri* induce secretogenic activity *ex vivo*. *PLoS ONE* **7**, e49980 (2012).
149. Ashida, H. *et al. Shigella* deploy multiple countermeasures against host innate immune responses. *Curr. Opin. Microbiol.* **14**, 16–23 (2011).
150. Sansonetti, P. J. *et al.* Caspase-1 activation of IL-1β and IL-18 are essential for *Shigella flexneri*-induced inflammation. *Immunity* **12**, 581–590 (2000).
151. Sansonetti, P. & Arondel, J. Interleukin-8 controls bacterial transepithelial translocation at the cost of epithelial destruction in experimental shigellosis. *Infect. Immun.* **67**, 1471–1480 (1999).
152. Kim, D. W. *et al.* The *Shigella flexneri* effector OspG interferes with innate immune responses by targeting ubiquitin-conjugating enzymes. *Proc. Natl Acad. Sci. USA* **102**, 14046–14051 (2005).
153. Arbibe, L. *et al.* An injected bacterial effector targets chromatin access for transcription factor NF-κB to alter transcription of host genes involved in immune responses. *Nat. Immunol.* **8**, 47–56 (2007).
154. Newton, H. J. *et al.* The type III effectors NleE and NleB from enteropathogenic *E. coli* and OspZ from *Shigella* block nuclear translocation of NF-κB p65. *PLoS Pathog.* **6**, e1000898 (2010).
155. Salgado-Pabón, W., Konradt, C., Sansonetti, P. J. & Phalipon, A. New insights into the crosstalk between *Shigella* and T lymphocytes. *Trends Microbiol.* **22**, 192–198 (2014).
156. Konradt, C. *et al.* The *Shigella flexneri* type three secretion system effector IpgD inhibits T cell migration by manipulating host phosphoinositide metabolism. *Cell Host Microbe* **9**, 263–272 (2011).
157. Salgado-Pabón, W. *et al. Shigella* impairs T lymphocyte dynamics *in vivo*. *Proc. Natl Acad. Sci. USA* **110**, 4458–4463 (2013).
158. Nothelfer, K. *et al.* B lymphocytes undergo TLR2-dependent apoptosis upon *Shigella* infection. *J. Exp. Med.* **211**, 1215–1229 (2014).
159. Sansonetti, P. J., Kopecko, D. J. & Formal, S. B. Involvement of a plasmid in the invasive ability of *Shigella flexneri*. *Infect. Immun.* **35**, 852–860 (1982).
160. Sansonetti, P. J., Kopecko, D. J. & Formal, S. B. *Shigella sonnei* plasmids: evidence that a large plasmid is necessary for virulence. *Infect. Immun.* **34**, 75–83 (1981).
This is the first study to demonstrate that the large plasmid is necessary for virulence in *Shigella* spp.
161. Rehel, N. & Szatmari, G. Characterization of the stable maintenance of the *Shigella flexneri* plasmid pHS-2. *Plasmid* **36**, 183–190 (1996).
162. Stieglitz, H. & Lipsky, P. Association between reactive arthritis and antecedent infection with *Shigella flexneri* carrying a 2-MD plasmid and encoding an HLA-B27 mimetic epitope. *Arthritis Rheum.* **36**, 1387–1391 (1993).
163. Gaston, J. *Shigella* induced reactive arthritis. *Ann. Rheum. Dis.* **64**, 517–518 (2005).
164. Hannu, T., Mattila, L., Siitonen, A. & Leirisalo-Repo, M. Reactive arthritis attributable to *Shigella* infection: a clinical and epidemiological nationwide study. *Ann. Rheum. Dis.* **64**, 594–598 (2005).
165. Calcuttawala, F., Hariharan, C., Pazhani, G. P., Ghosh, S. & Ramamurthy, T. Activity spectrum of colicins produced by *Shigella sonnei* and genetic mechanism of colicin resistance in conspecific *S. sonnei* strains and *Escherichia coli*. *Antimicrob. Agents Chemother.* **59**, 152–158 (2015).
166. The, H. C. *et al.* The introduction and establishment of fluoroquinolone resistant *Shigella sonnei* into Bhutan. *Microb. Genom.* http://dx.doi.org/10.1099/mgen.0.000042 (2015).

© 2016 Macmillan Publishers Limited. All rights reserved.

167. Ferreccio, C. *et al.* Epidemiologic patterns of acute diarrhea and endemic *Shigella* infections in children in a poor periurban setting in Santiago, Chile. *Am. J. Epidemiol.* **134**, 614–627 (1991).
168. Noriega, F. R. *et al.* Strategy for cross-protection among *Shigella flexneri* serotypes. *Infect. Immun.* **67**, 782–788 (1999).
169. Feil, E. J. The emergence and spread of dysentery. *Nat. Genet.* **44**, 964–965 (2012).
170. Levine, M., Kotloff, K. L, Barry, E. M., Pasetti, M. F. & Sztein, M. B. Clinical trials of *Shigella* vaccines: two steps forward and one step back on a long, hard road. *Nat. Rev. Microbiol.* **5**, 540–553 (2007).
171. Barry, E., Pasetti, M. & Sztein, M. Progress and pitfalls in *Shigella* vaccine research. *Nat. Rev. Gastroenterol. Hepatol.* **10**, 245–255 (2013).
172. Walker, R. I. An assessment of enterotoxigenic *Escherichia coli* and *Shigella* vaccine candidates for infants and children. *Vaccine* **33**, 954–965 (2015).
A comprehensive and up-to-date review evaluating the current approaches for a *Shigella* spp. vaccine.
173. Martinez-Becerra, F. J. *et al.* Characterization of a novel fusion protein from IpaB and IpaD of *Shigella* spp. and its potential as a pan-*Shigella* vaccine. *Infect. Immun.* **81**, 4470–4477 (2013).
174. Riddle, M. S. *et al.* Safety and immunogenicity of an intranasal *Shigella flexneri* 2a Invaplex 50 vaccine. *Vaccine* **29**, 7009–7019 (2011).

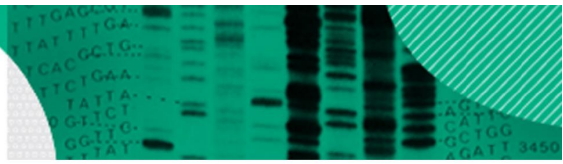
Acknowledgements
H.C.T., D.P.T. and N.R.T. are funded by the Wellcome Trust (grant 098051); K.E.H is funded by the Australian National Health and Medical Research Council (NHMRC); and S.B. is a Sir Henry Dale Fellow, funded by the Wellcome Trust and the Royal Society (grant 100087/Z/12/Z).

Competing interests statement
The authors declare no competing interests.

FURTHER INFORMATION
WHO: http://www.who.int/
GEMS: http://medschool.umaryland.edu/GEMS/
ALL LINKS ARE ACTIVE IN THE ONLINE PDF

© 2016 Macmillan Publishers Limited. All rights reserved.

MICROBIAL GENOMICS
Bases to Biology



Research

Introduction and establishment of fluoroquinolone-resistant *Shigella sonnei* into Bhutan

Hao Chung The,[1] Maia A. Rabaa,[1,2] Duy Pham Thanh,[1] Sirigade Ruekit,[3] Sonam Wangchuk,[4] Tshering Dorji,[4] Kinzang Pem Tshering,[5] To Nguyen Thi Nguyen,[1] Phat Voong Vinh,[1] Tuyen Ha Thanh,[1] Chau Nguyen Ngoc Minh,[1] Paul Turner,[2,6] Poda Sar,[6] Guy Thwaites,[1,2] Kathryn E. Holt,[7] Nicholas R. Thomson,[8,9] Ladaporn Bodhidatta,[3] Carl Jeffries Mason[3] and Stephen Baker[1,2,9]

[1]The Hospital for Tropical Diseases, OUCRU, Ho Chi Minh City, Vietnam

[2]Centre for Tropical Medicine, Oxford University, Oxford, UK

[3]Department of Enteric Diseases, AFRIMS, Bangkok, Thailand

[4]Public Health Laboratory, Department of Public Health, Ministry of Health, Royal Government of Bhutan, Thimphu, Bhutan

[5]Department of Pediatrics, Jigme Dorji Wangchuk Referral Hospital, Kawa Jangsa, Thimphu, Bhutan

[6]Cambodia–Oxford Medical Research Unit, Angkor Hospital for Children, Siem Reap, Cambodia

[7]Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia

[8]The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK

[9]The London School of Hygiene and Tropical Medicine, London, UK

Correspondence: Stephen Baker (sbaker@oucru.org)

DOI: 10.1099/mgen.0.000042

Shigella sonnei is a major contributor to the global burden of diarrhoeal disease, generally associated with dysenteric diarrhoea in developed countries but also emerging in developing countries. The reason for the recent success of *S. sonnei* is unknown, but is likely catalysed by its ability to acquire resistance against multiple antimicrobials. Between 2011 and 2013, *S. sonnei* exhibiting resistance to fluoroquinolones, the first-line treatment recommended for shigellosis, emerged in Bhutan. Aiming to reconstruct the introduction and establishment of fluoroquinolone-resistant *S. sonnei* populations in Bhutan, we performed whole-genome sequencing on 71 *S. sonnei* samples isolated in Bhutan between 2011 and 2013.We found that these strains represented an expansion of a clade within the previously described lineage III, found specifically in Central Asia. Temporal phylogenetic reconstruction demonstrated that all of the sequenced Bhutanese *S. sonnei* diverged from a single ancestor that was introduced into Bhutan around 2006. Our data additionally predicted that fluoroquinolone resistance, conferred by mutations in *gyrA* and *parC*, arose prior to the introduction of the founder strain into Bhutan. Once established in Bhutan, these *S. sonnei* had access to a broad gene pool, as indicated by the acquisition of extended-spectrum β-lactamase-encoding plasmids and genes encoding type IV pili. The data presented here outline a model for the introduction and maintenance of fluoroquinolone-resistant *S. sonnei* in a new setting. Given the current circulation of fluoroquinolone-resistant *S. sonnei* in Asia, we speculate that this pattern of introduction is being recapitulated across the region and beyond.

Received 8 September 2015; Accepted 19 November 2015

© 2015 The Authors. Published by Microbiology Society

Downloaded from www.microbiologyresearch.org by
IP: 118.69.34.244
On: Mon, 28 Dec 2015 07:55:31

Keywords: Bhutan; fluoroquinolone; resistance; *Shigella sonnei*.

Abbreviations: ESBL, extended-spectrum β-lactamase; HPD, highest posterior density; JDWNRH, Jigme Dorji Wangchuk National Referral Hospital; MCMC, Markov Chain Monte Carlo; MDR, multidrug-resistant; MRCA, most recent common ancestor; WGS, whole-genome sequencing.

Data statement: Whole-genome sequence reads of *Shigella sonnei* used in this study have been deposited to the European Nucleotide Archive. Table S1 has been deposited in FigShare. All supporting data, code and protocols have been provided within the article or through supplementary data files.

Data Summary

1. Whole-genome sequence reads of *Shigella sonnei* used in this study have been deposited to the European Nucleotide Archive. The run accession numbers and related metadata are detailed in Table S1 which has been deposited in FigShare: http://figshare.com/articles/The_introduction_and_establish ment_of_fluoroquinolone_resistant_*Shigella*_sonnei_ into_Bhutan/1610693

Impact Statement

The bacterium *Shigella sonnei* is a major global cause of diarrhoeal disease. This organism has generally been associated with infections in developed countries, but is currently dramatically expanding into industrializing countries. The mechanisms underpinning the spread of *S. sonnei* are not precisely understood, but we hypothesize that their spread is partly facilitated by resistance to antimicrobials, including the current therapeutic choice to combat *Shigella*, the fluoroquinolones. To understand how the fluoroquinolone-resistant *S. sonnei* may become established in a new location, we decoded the total DNA sequences of 71 *S. sonnei* samples isolated in Bhutan in South Asia between 2011 and 2013. We found that these strains were likely descended from a single isolate that was introduced into Bhutan around 2006. Furthermore, we additionally found that fluoroquinolone resistance emerged prior to the introduction of this founder *S. sonnei* into Bhutan and these strains then became established, gaining additional genes and resistance to other antimicrobials. Our data describe how drug-resistant *S. sonnei* may become resident in new locations, and we predict that a similar situation is occurring currently with the epidemic of fluoroquinolone-resistant *S. sonnei* in Asia and beyond.

Introduction

It has recently been estimated that the global burden of diarrhoeal disease in children <5 years of age is $\sim$1.7 billion new infections with 0.7 million deaths per year (Walker *et al.*, 2013). These figures make diarrhoeal diseases, after respiratory tract infections, the second most common cause of mortality in young children in low-income countries (Liu *et al.*, 2012; WHO, 2014). Further contemporary data from the Global Enteric Multicentre Study, the largest prospective case–control study on paediatric diarrhoeal illnesses in sub-Saharan Africa and South Asia ever conducted, identified *Shigella* to be amongst the most common diarrhoeal pathogens in these high-incidence settings (Kotloff *et al.*, 2012, 2013).

The members of the genus *Shigella* are Gram-negative pathogens that are closely related to *Escherichia coli* (Pupo *et al.*, 2000) and cause dysenteric diarrhoea (stools containing blood and/or mucus) via a cascade of virulence factors encoded principally on a large signature virulence plasmid (Sansonetti *et al.*, 1982). The genus contains four species: *Shigella sonnei*, *Shigella flexneri*, *Shigella boydii* and *Shigella dysenteriae*, with *S. sonnei* and *S. flexneri* dominating in developed and developing countries, respectively. However, this pattern is changing, as *S. sonnei* is currently replacing *S. flexneri* in countries undergoing economic transition (Vinh *et al.*, 2009; Thompson *et al.*, 2015). This species replacement has positive and negative ramifications. By replacing *S. flexneri*, which is very diverse with multiple serotypes with respect to the single serotype *S. sonnei* (Connor *et al.*, 2015), as the leading cause of dysentery means that it may be possible to control disease through vaccination (Levine *et al.*, 2007). However, *S. sonnei* is also highly adept at acquiring antimicrobial resistance genes and resistance-associated mutations (Huang *et al.*, 2005; Seol *et al.*, 2006; Nguyen *et al.*, 2010), limiting treatment options and making control through antimicrobial treatment alone improbable.

Recent advances in whole-genome sequencing (WGS) and analytical methodologies have enhanced our understanding of the epidemiology, international spread and antimicrobial susceptibility trends of *S. sonnei* (Holt *et al.*, 2012, 2013). We now know that *S. sonnei* is formed of three main lineages (I, II and III), which are all related to a common ancestor that arose in Europe $\sim$400 years ago (Holt *et al.*, 2012). Lineage III, which is multidrug-resistant (MDR), now dominates globally (Holt *et al.*, 2012). We have also previously shown that once introduced into a new geographical region, into an apparently naive human population, *S. sonnei* can establish new isolated clonal populations which evolve locally (Holt *et al.*, 2013). Antimicrobial resistance is a key driver of the recent *S. sonnei* global expansions marked by the acquisition of antimicrobial resistance genes and/or mutations in genes encoding the target proteins for antimicrobials (Pu *et al.*, 2015).

Downloaded from www.microbiologyresearch.org by
IP: 118.69.34.244
On: Mon, 28 Dec 2015 07:55:31

Bhutan is a lower middle-income country in South Asia, sandwiched between China and the northwest Indian states of Assam and West Bengal. In close proximity to Nepal (<70 km of Indian territory separates the two countries), Bhutan is a Himalayan nation with an estimated population of 750 000 people, of which the majority (69 %) live in remote rural locations (Office of the Census Commissioner – Royal Government of Bhutan, 2006). Like other low-income countries in the region, diarrhoeal disease is a major public health problem in Bhutan, and both *S. flexneri* and *S. sonnei* are endemic (Marfin *et al.*, 1994; Ruekit *et al.*, 2014). Between March 2011 and October 2013, a diarrhoeal disease surveillance study was conducted in the Bhutanese capital, Thimphu. During this study, there was an increased number of *S. sonnei* isolated between June and July. Antimicrobial susceptibility testing, PFGE and plasmid replicon typing of 29 of these *S. sonnei* isolates showed that all were identical by PFGE pattern and antimicrobial susceptibility (Ruekit *et al.*, 2014). All these *S. sonnei* isolates were resistant to ciprofloxacin (MIC$\geq$3 μg μl^{-1}), which was infrequently observed in strains circulating globally during the same period (Ruekit *et al.*, 2014). This study suggested that these *S. sonnei* isolates were closely related and may have been a recent introduction into Bhutan.

To test the hypothesis that these fluoroquinolone-resistant *S. sonnei* had recently been introduced to Bhutan and to further characterize these Bhutanese isolates in the context of global *S. sonnei*, we performed WGS on 71 *S. sonnei* isolated in Bhutan between 2011 and 2013.

Methods

Sample collection. The Research Ethics Board of Health in Bhutan and the Institutional Review Board of the Walter Reed Army Institute of Research provided ethical approval for the study. Stool samples were collected as described previously (Ruekit *et al.*, 2014). Stool samples were collected from cases and controls, and cultured on MacConkey and Hektoen agar, and incubated overnight at 37 °C. Non-lactose-fermenting colonies were confirmed to be *Shigella* by standard biochemical testing; serotype was identified by O-antigen agglutination (Denka Seiken).

Stool samples were collected during a diarrhoea surveillance study conducted at the Jigme Dorji Wangchuk National Referral Hospital (JDWNRH) in Thimphu, Bhutan, between March 2011 and October 2013, as described previously (Ruekit *et al.*, 2014). Briefly, children aged between 3 months and 5 years presenting to JDWNRH with acute diarrhoea (three or more loose stools in the previous 24 h, with symptoms lasting $\leq$72 h at time of recruitment) and those with no diarrhoeal symptoms in the previous 2 weeks were eligible for recruitment as cases and controls, respectively (informed consent permitting).

All *Shigella* isolates were subjected to antimicrobial susceptibility testing using the disk diffusion method as recommended by Clinical and Laboratory Standards Institute guidelines (CLSI, 2012). The following antimicrobials were selected for susceptibility testing: azithromycin, nalidixic acid, ciprofloxacin, ampicillin, trimethoprim/sulfamethoxazole, ceftriaxone, streptomycin and tetracycline.

WGS. During the surveillance study a total of 112 *Shigella* were isolated (74 *S. sonnei*, 30 *S. flexneri*, seven *S. boydii* and one *S. dysenteriae*); the 74 *S. sonnei* were selected for WGS. The majority (64/74) of the selected *S. sonnei* were isolated from symptomatic diarrhoeal stool samples with the remaining strains isolated from asymptomatic infections.

Genomic DNA was extracted from all *S. sonnei* isolates using a Wizard Genomic DNA Extraction kit (Promega) according to the manufacturer's recommendations. For each sample, 2 μg genomic DNA was subjected to WGS on an Illumina HiSeq2000 platform, following the manufacturer's protocols to generate 100 bp paired-end reads (Quail *et al.*, 2008). All reads were mapped to the reference sequence of *S. sonnei* Ss046 (GenBank accession number NC_007384) using SMALT (version 0.7.4). Candidate SNPs were called against the reference sequence and filtered using SAMtools (Li *et al.*, 2009). Low-quality SNPs were removed according to the following criteria: consensus quality <50, mapping quality <30, ratio of SNPs to reads at a position <75 %, read depth <4, number of reads per strand <2, strand bias <0.001, mapping bias <0.001 or tail bias <0.001. Three samples failed due to insufficient read depth, resulting in a final sample set including 71 *S. sonnei* whole-genome sequences from Bhutan.

Phylogenetic analysis, temporal analysis and Bayesian phylogenetic inference. In order to place the Bhutanese isolates in the context of global diversity, we aligned the Bhutanese *S. sonnei* sequences with publicly available sequences utilized in a previous global analysis (Holt *et al.*, 2012). An initial phylogeny was inferred with these data using RAxML under the GTRGAMMA model. To further assess the relationship of the Bhutanese isolates within Central Asia, the previous alignment was then subsampled to include only the 71 Bhutanese sequences along with 20 other closely related Central Asia sequences. Previously described mobile genetic elements (Holt *et al.*, 2012) and potential recombinant regions were removed from the alignment using Gubbins (Croucher *et al.*, 2014). Further removal of gaps, indeterminate bases, invariant sites and 17 Bhutanese sequences that were identical to others in the dataset resulted in a final alignment of 996 SNPs across 74 taxa. A maximum-likelihood phylogeny was inferred from these sequences using GARLI (Zwickl, 2006) with 1000 bootstrap replicates under the TVM (transversion) model of nucleotide substitution, which was determined to be the best-fit model using jModelTest (Posada, 2008).

Downloaded from www.microbiologyresearch.org by
IP: 118.69.34.244
On: Mon, 28 Dec 2015 07:55:31

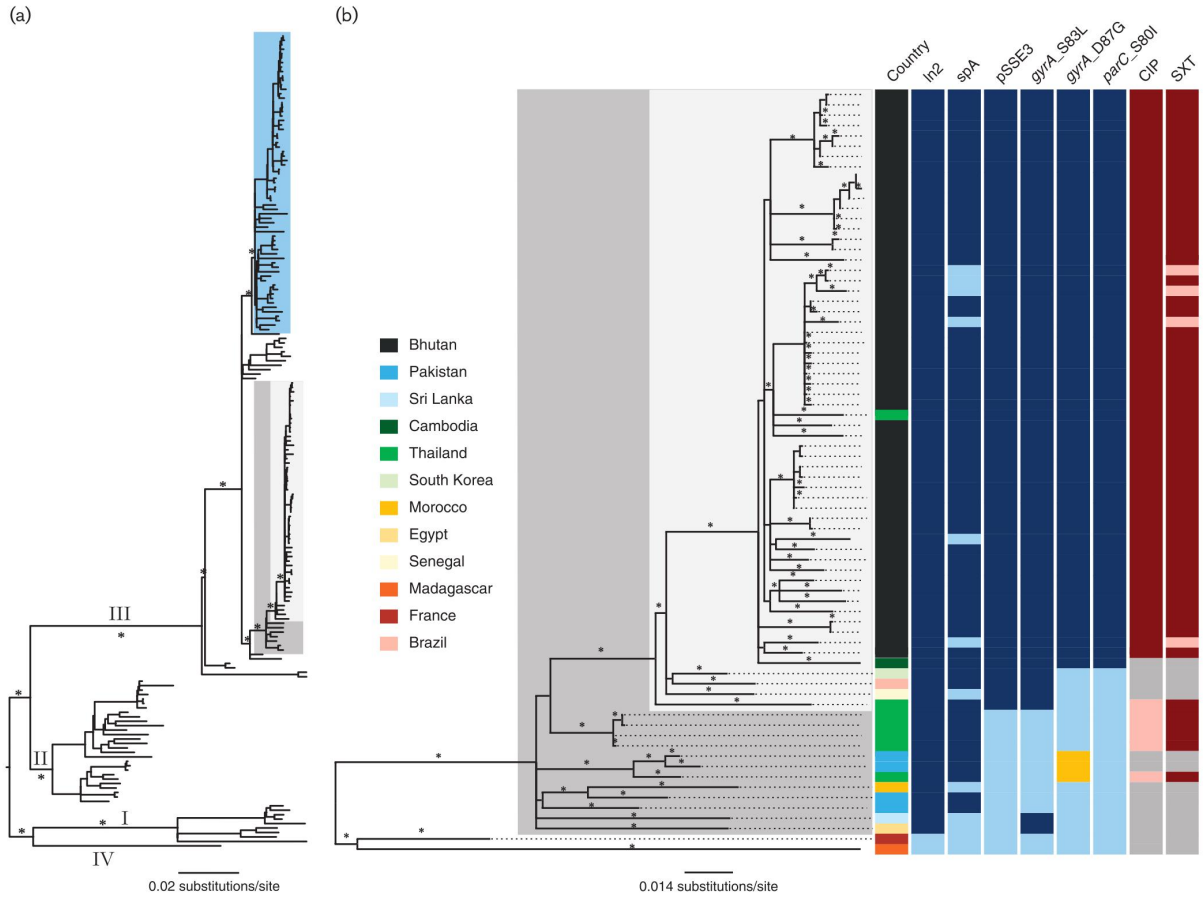


Fig. 1. The phylogenetic structure of Bhutanese *S. sonnei* in the context of the global phylogeny and Central Asia clade. (A) Midpoint rooted maximum-likelihood phylogenetic tree of 183 *S. sonnei* strains (135 from global collection and 48 from Bhutan) reconstructed using 5393 SNPs; asterisks indicate bootstrap support values ≥98 % on major branches. Numbers above major branches represent ones leading to major lineages (I, II, III and IV). The light blue box highlights the Global III clade; the dark grey box overlaid on the tree highlights strains belonging to the Central Asia clade; the smaller, light grey box highlights the primarily Bhutanese Central Asia clade. (B) Magnified view of the maximum-likelihood phylogenetic tree of the Central Asia clade, including 74 *S. sonnei* strains (54 from Bhutan and 20 others for phylogenetic context), reconstructed using 996 SNPs; asterisks indicate bootstrap support values ≥80 %. The tree is midpoint rooted for purposes of clarity. Columns to the right of the phylogenetic tree show isolate metadata including: country of isolation (colours indicated in the key); the presence (dark blue) or absence (light blue) of specific genetic elements and mutations [In2, spA, pSSE3, *gyrA* (S83L), *gyrA* (D87G), *parC*(S80I)], or the presence of an alternative mutation (yellow) [*gyrA* (D87Y)]; and resistance profiles against ciprofloxacin (CIP) and trimethoprim/sulfamethoxazole (SXT), where resistance is indicated in dark red, susceptibility in light red and missing data in grey.

In order to thoroughly examine the population and temporal structure of Bhutanese *S. sonnei* and the population from which they emerged, all sequences from Bhutan, including identical sequences removed in the previous phylogeny, were aligned with the six most closely related isolates from other countries and trimmed down to include only the 544 SNPs present in the alignment of 77 taxa. These data were then utilized for maximum-likelihood reconstruction using PhyML under the TVM + Γ_4 substitution model (Guindon *et al.*, 2010). In order to investigate whether temporal signal was embedded in the resulting maximum-likelihood phylogeny, we applied Path-O-Gen (version 1.4) to assess the linear relationship between root-to-tip divergence and date of isolation (in day, month, year).

Bayesian MCMC (Markov Chain Monte Carlo) deployed in BEAST (version 1.8.0) (Drummond & Rambaut, 2007)

Downloaded from www.microbiologyresearch.org by
IP: 118.69.34.244
On: Mon, 28 Dec 2015 07:55:31

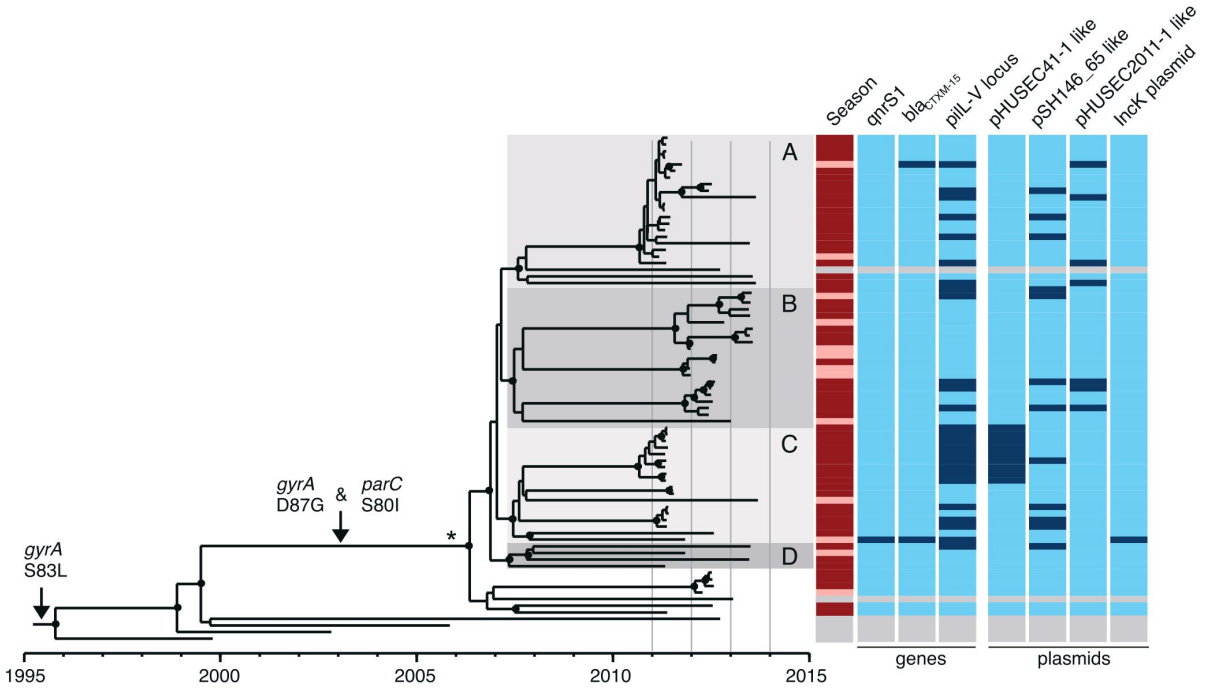


Fig. 2. Temporal phylogenetic reconstruction of Bhutanese *Shigella sonnei* between 2011 and 2013.Image shows a maximum clade credibility phylogenetic reconstruction of *S. sonnei* isolated primarily in Bhutan over the study period. Distinct subpopulations of Bhutanese *S. sonnei* are highlighted and indicated by letters. Black circles indicate posterior probability support ≥ 80 % on internal nodes. The asterisk indicates the branch leading to the Bhutanese *S. sonnei* clade, which represents 22 lineage-defining SNPs. Arrows denote branches characterized by select substitution events. Columns to the right of the phylogeny show metadata including: season (dark red, monsoon season: June–September; light red, other), and presence (dark blue) or absence (light blue) of mobile genetic elements, including specific genes (*qnrS1*, $bla_{\text{CTXM-15}}$, *pilL–V* locus) and plasmids (pHUSEC41-1-like, pSH146_65-like, pHUSEC2011-1-like, IncK). Grey bars in metadata columns indicate sequences from *S. sonnei* isolated outside of Bhutan.

was then used to infer the evolutionary dynamics of the aforementioned dataset, with a particular focus on Bhutanese *S. sonnei*. To determine the best-fit priors to the data at hand, multiple BEAST runs were conducted using the TVM or GTR nucleotide substitution model with constant, exponential growth or Bayesian skyline demographic history priors, in combination with either a strict or a relaxed molecular clock (uncorrelated lognormal distribution) (Drummond & Rambaut, 2007). For each BEAST analysis, marginal likelihood estimation was conducted using both path sampling and stepping-stone sampling approaches to facilitate model selection (Baele *et al.*, 2012, 2013). Bayes factor comparison indicated the TVM$+\Gamma_4$ substitution model with a relaxed log-normal molecular clock and a piecewise Bayesian skyline demographic model as the best-fit model to the data. Analyses using these priors were then performed in triplicate using a continuous 100 million generation MCMC chain with samples taken every 10 000 generations and parameter convergence (indicated by effective sample size values >200) was assessed in Tracer (version 1.5). LogCombiner (version 1.8) was used to combine triplicate runs, with removal of 20 % burn-in.

Gene content analysis. *De novo* assembly for each read set was performed using the assembler Velvet (version 1.2.03) and VelvetOptimizer, with each read set mapped back to each assembly (Zerbino & Birney, 2008). Contigs <300 bp were discarded. Assemblies were annotated using the rapid prokaryotic genome annotation tool Prokka (Seemann, 2014). To identify contigs associated with the accessory genome, contigs were first-ordered with *S. sonnei* Ss046 and its large virulence plasmid pSS046 as references using ABACAS. The remaining contigs were visualized and examined in Artemis. Presence of

Downloaded from www.microbiologyresearch.org by
IP: 118.69.34.244
On: Mon, 28 Dec 2015 07:55:31

Table 1. Non-synonymous mutations separating the MRCA of all Bhutanese *S. sonnei* from other Central Asia clade strains

Gene	Mutation	Function	SIFT score*/effect
cusS	G343S	Sensory histidine kinase of CusSR, regulating the expression of the CusCFBA operon to confer resistance to high copper and silver concentrations	0.77
ycgB	Y125H	Unknown function	0
mppA	G344R	Periplasmic binding component of murein tripeptide in oligopeptide transport	0
hisC	P98S	Histidinol-phosphate aminotransferase, catalysing the biosynthesis of histidine	0.03
eutL	A22S	Carboxysomal shell protein, structural component of proteinaceous microcompartment for ethanolamine ammonia lyase	0.5
hycC	G538D	Component of HycBCDEFG, hydrogenase component (hydrogenase 3) of formate hydrogenlyase; HycC encodes an extremely hydrophobic protein with homology to NADH : ubiquinone oxidoreductase	0
sdaB	A407S	L-Serine deaminase II, catalysing the conversion of serine into pyruvate and ammonia	0.33
metC	A2T	Cystathionine B-lyase and cysteine desulfhydrase, catalysing the conversion cystathionine to homocysteine	0.7
rbsK	A50T	Ribokinase responsible for metabolism of D-ribose	0
gyrA	D87G	DNA gyrase, subunit A	*Known effect*: resistance to fluoroquinolones
parC	S80I	DNA topoisomerase IV subunit A	*Known effect*: resistance to fluoroquinolones

* SIFT scores ≤0.05 indicate a significant effect on protein function based on SIFT analysis.

specific accessory genetic determinants was confirmed by BLAST and visualized using ACT (Carver *et al.*, 2008).

For each isolate, the entire resistance gene content (resistome) was identified by mapping the assembly to a manually curated resistome database as described previously (Chung The *et al.*, 2015). MLST was performed for all *Shigella* isolates by profiling the seven MLST genes (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, *recA*). Mutations in specific genes were manually assessed by retrieving their alignments from the annotations and the effects of non-synonymous mutations were predicted using the SIFT web server (http://sift.jcvi.org/) (Sim *et al.*, 2012). An *in silico* plasmid incompatibility-type PCR was performed using an in-house script and bespoke primer set.

Results

Bhutanese *S. sonnei* lies within lineage III (Central Asia clade)

All *S. sonnei* (n=74) isolated during this study were subjected to WGS; 71 isolates produced sufficient sequence for analysis. To contextualize the Bhutanese *S. sonnei* within the known global *S. sonnei* population structure, we constructed a phylogeny that included the Bhutanese sequences and those from a previous global *S. sonnei* sequencing project (Holt *et al.*, 2012); within this phylogeny, all Bhutanese *S. sonnei* fell into a single clade within lineage III (Central Asia clade), which is a sister clade to the intercontinental expanding Global III clade (Fig. 1A). To investigate the phylogenetic relationships within the Central Asia clade, we concatenated 74 of the non-duplicate whole-genome sequences, including 54 from Bhutan and 20 selected from other countries. A maximum-likelihood phylogeny reconstructed from this alignment showed a close clonal relationship between the Bhutanese *S. sonnei*. Showing a similarly close relationship to the Bhutanese isolates, and interspersed within this clade, were also two isolates from Thailand and Cambodia, likely representing importations following regional travel (Fig. 1B).

Temporal phylogenetic reconstruction of *S. sonnei* in Bhutan

As there was such a strong temporal signature for the number of observed mutations in the global *S. sonnei* population (Holt *et al.*, 2013), we explored the temporal structure of the *S. sonnei* isolated in Bhutan using Bayesian phylogenetic methods. To do this, the nucleotide substitution rate of the primarily Bhutanese lineage of Central

Downloaded from www.microbiologyresearch.org by
IP: 118.69.34.244
On: Mon, 28 Dec 2015 07:55:31

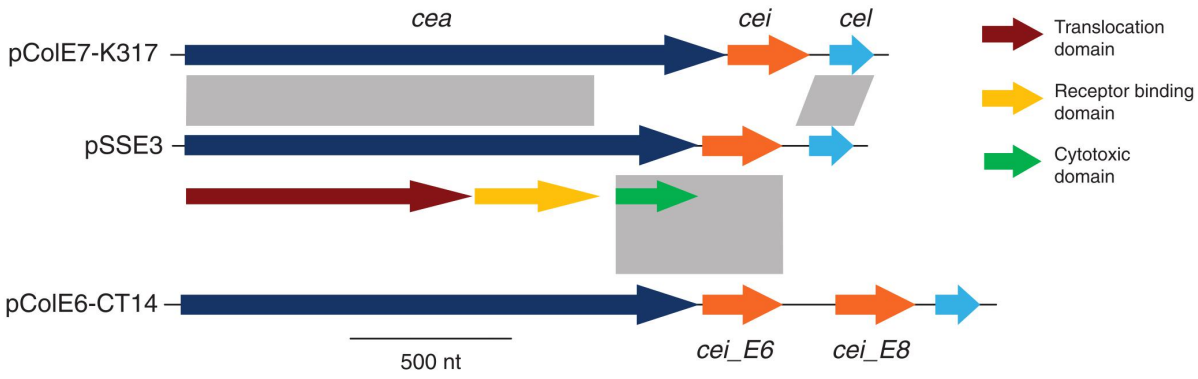


Fig. 3. Novel recombinant hybrid colicin biosynthesis cluster encoded on pSSE3.Schematic representation of the colicin biosynthesis cluster on plasmid pSSE3 in comparison with colicin gene clusters of pColE7-K317 and pColE6-CT14. Grey vertical bars show homology in DNA sequences (with identity >80 %). The colicin structural gene (*cea*), immunity gene (*cei*) and lysis gene (*cel*) are annotated as shown. The three domains of the colicin structural gene are shown the key.

Asia was determined to be 7.34×10^{-7} substitutions per site per year [95 % highest posterior density (HPD): 5.87×10^{-7} to 8.83×10^{-7}], which is comparable to the previously calculated substitution rate of *S. sonnei* lineage III (Holt *et al.*, 2013). Using these data, we estimate that the most recent common ancestor (MRCA) of the Bhutanese *S. sonnei* existed around 2006 (95 % HPD 2005–2007), with the majority diverging from a hypothetical MRCA around 2007 (95 % HPD 2005–2008) (Fig. 2). The results of this analysis were further supported by the linear relationship between root-to-tip branch length and date of isolation (R^2=0.8412). Since its projected introduction in 2006, this *S. sonnei* lineage diverged into at least four sublineages (labelled A, B, C and D in Fig. 2), which have been transmitted endemically within the Bhutanese capital, Thimphu: isolates taken from cases in June/July 2011 (25/71 isolates) were found to belong to two subpopulations (A and C). Sublineage B, which was not represented in June/July 2011, dominated in the subsequent years (Fig. 2).

Mutations specific to Bhutanese *S. sonnei*

To identify potential genetic mechanisms associated with diversification of the Bhutanese *S. sonnei*, we compared their content of unique SNPs in comparison with the three most closely related outgroup isolates, taken from global isolates belonging to the same global lineage but collected outside of Bhutan (Fig. 2). The Bhutanese isolates were distinguished from these by the presence of 22 lineage-specific consensus SNPs. Of these 22 SNPs, five fell in intergenic regions and 17 mapped within predicted coding sequences; 11 represented non-synonymous and six were synonymous mutations. The genes bearing non-synonymous mutations are summarized in Table 1 together with their predicted impact on function. These include mutations in *gyrA* and *parC* associated with antimicrobial resistance, as well as those located in *ycgB*, *mppA*, *hisC* and *rbsK*, as determined by *in silico* SIFT (see Methods) analysis. With the exception of *ycgB* encoding a protein with unknown function, these genes all encode proteins predicted to be involved in the metabolism of histidine, ribose and formate, as well as transport of murein tripeptide (Table 1).

Accessory genome and impact of antimicrobial resistance genes in Bhutanese *S. sonnei*

The composition of the accessory genome can be used to indicate the capacity of the host species to sample the local gene pool and also to highlight themes in the patterns of acquisition that may be under selection. We performed a comparative gene content analysis of Bhutanese and non-Bhutanese *S. sonnei* belonging to the Central Asia clade. These data revealed events that appear to have been pivotal in evolution and spread of the Central Asia clade. Notably, all the isolates of this clade possessed a specific class 2 integron In2 (*dfrA1*, *sat1*), conferring resistance to trimethoprim and streptothricin. These strains additionally gained the *S. sonnei* spA plasmid carrying *sul2*, *strAB* and *tetRA* genes, which further expanded the spectrum of antimicrobial resistance to incorporate sulphonamides, streptomycin and tetracycline resistance, respectively (Fig. 1B). This resistance gene profile concurs with the observation made in the previous global *S. sonnei* study (Holt *et al.*, 2012). It is of note that the class 2 integron of the Central Asia clade differs from that of Global III in the absence of *aadA1*, which confers resistance to aminoglycosides. Moreover, it was clear that isolates within this clade have, over time, accumulated chromosomal mutations conferring fluoroquinolone resistance (Fig. 1B). From the lineage-specific phylogeny, we predicted that the first mutation in *gyrA* (S83L) occurred before 1996, and sup-

Downloaded from www.microbiologyresearch.org by
IP: 118.69.34.244
On: Mon, 28 Dec 2015 07:55:31

plementary mutations in *parC* (S80I) and secondary mutations in *gyrA* (D87G), leading to high-level fluoroquinolone resistance, were fixed in the Bhutanese clade prior to its expansion and diversification in 2006 (Fig. 2).

A further pivotal event in the emergence of this Central Asia clade prior to its introduction into Bhutan was the fixation of the previously described E-type colicin plasmid pSSE3 (with 100 % identity in DNA sequence). This E-type colicin has been shown to be antagonistic against a variety of diarrhoeogenic *E. coli*, but not against commensal *E. coli* (Calcuttawala *et al.*, 2015). Plasmid pSSE3, which was found in all Bhutanese isolates (Fig. 1B), also carries an immunity function that protects the host cell against the activity of the corresponding colicin. Although it was previously characterized as a ColE3 plasmid by PCR, sequence analysis of pSSE3 indicated that the colicin biosynthesis cluster was a novel recombinant hybrid with an undetermined spectrum of activity (Fig. 3). On closer examination, the first and second domains (translocation and receptor-binding) of the colicin structural gene (*cea*) and the lysis gene (*cel*) exhibited substantial amino acid sequence identity (98 %) to that of colicin E7 (ColE7). The third domain (cytotoxic) encoded by *cea* was most closely related to that of ColE6 (94 % protein sequence identity), predicting that the colicin's target is actually 16S rRNA rather than DNA, which is the target of the ColE7 counterpart (Kleanthous, 2010). However, the product of the colicin immunity gene (*cei*) exhibited only 84 % protein sequence identity to that of colicin E6.

Genetic material that was acquired sporadically across the Central Asia clade within the Bhutanese isolates included the extended-spectrum β-lactamase (ESBL) gene $bla_{\text{CTXM-15}}$, which was independently acquired by two *S. sonnei* on different incompatibility plasmids (IncK and IncI1), conferring resistance to third-generation cephalosporins (Fig. 2). The plasmid replication and transfer region of the IncK plasmid shared substantial sequence homology with that of the enteroaggregative *E. coli* IncI1 plasmid pSERB1. This plasmid in the Bhutanese *S. sonnei* incorporated the type IV pilus biosynthesis cluster *pilL–V* (Dudley *et al.*, 2006). For the *S. sonnei* containing the IncK plasmid, the $bla_{\text{CTXM-15}}$ gene was encoded along with the fluoroquinolone resistance enhancement gene *qnrS1* on an IS*26* transposon. The *S. sonnei* IncI1 plasmid (denoted pHUSEC2011-1-like) was almost identical ($\geq$99 % DNA sequence identity) to the replication, conjugation and type IV pilus biosynthesis cluster *pilL–V* in pHUSEC2011-1. The pHUSEC2011-1-like plasmid was also independently acquired by six additional strains, although in these cases the $bla_{\text{CTXM-15}}$ gene was not co-transferred (Fig. 2).

Lastly, nine *S. sonnei* strains isolated from cases occurring in June/July 2011 acquired an IncB-O plasmid, sharing homology with the MDR plasmid pHUSEC41-1 originally identified in an older O104 : H4 enteroaggregative Shiga toxin-producing *E. coli* (HUSEC41), which included a type IV pilus biosynthesis cluster *pilL–V* identical to that found on plasmid pSERB1 (Künne *et al.*, 2012). We further identified 11 *S. sonnei* strains harbouring a plasmid with homology to *Salmonella enterica* plasmid pSH146_65, additionally carrying a distinct type IV pilus cluster *pilL–V* (Han *et al.*, 2012).

Discussion

Here, we have used WGS to define a population of *S. sonnei* causing disease in the South Asian land-locked nation of Bhutan. These strains belong to an expansion of lineage III, which we have named the Central Asia clade, which was introduced into Bhutan around 2006. To the best of our knowledge, this is the first time that WGS has been used on this scale to study a population of *S. sonnei* in South Asia, a region with a substantial burden of diarrhoeal disease. In contrast to recent observations from Vietnam, the sequenced *S. sonnei* in Bhutan demonstrated no evidence of a single clonal expansion. This is perhaps due to the short sampling time frame or other factors such as the lack of a definitive selective sweep event. However, the evolution history of Central Asia clade *S. sonnei* resembles the two initial sweep events observed in Vietnam [associated with fixation of a colicin plasmid and mutation(s) in *gyrA*] (Holt *et al.*, 2013), showing that the two geographically and phylogenically unrelated *S. sonnei* populations followed a similar evolutionary trajectory. After its introduction to Bhutan, our data predict that these *S. sonnei* had access to a substantial gene pool, as indicated by the sporadic acquisition of ESBL plasmids and type IV pili. The gains of ESBL plasmids, although infrequent, could potentially become the next selective sweep for the Bhutan's *S. sonnei* population as observed in Vietnam, given the selection path toward resistance is maintained. The characteristically thin type IV pilus has been shown to improve the efficiency and specificity of conjugation in liquid environments (Kim & Komano, 1997), and also facilitates adherence, interbacterial contact and twitching motility (Berry & Pelicic, 2015). Although this structure is known to be widespread in prokaryotes, its impact on the lifestyle of *Shigella* is not well described and requires further experimental investigation (Holt *et al.*, 2013; Rohmer *et al.*, 2014).

In addition, the Bhutanese *S. sonnei* population is characterized by the presence of numerous non-synonymous mutations, most of which are found in metabolic pathways. The genus *Shigella* is well known for its adaptive intracellular lifestyle, enabling its shedding of redundant or unnecessary functions during the course of evolution (Prosseda *et al.*, 2012). Our *in silico* analysis predicted that some mutations could affect metabolic activities, which could lead to degradation. In particular, the affected D-ribose utilization and formate fermentation appear to be redundant as energy production in *Shigella* predominantly occurs via the uptake and metabolism of pyruvate into acetate (Kentner *et al.*, 2014). Biosynthesis of histidine is also deemed inessential as the bacterium could readily transport it from the host's cytosol. However, the true

Downloaded from www.microbiologyresearch.org by
IP: 118.69.34.244
On: Mon, 28 Dec 2015 07:55:31

impacts of the aforementioned mutations should be confirmed by functional analyses *in vitro* in future studies.

We were able to use WGS to identify the major events during the evolution of this lineage prior to, and after, its introduction into a new setting. Our data imply that the acquisition of a class 2 integron and a spA plasmid, together with the rapid accumulation of mutations associated with fluoroquinolone resistance, rendered the Bhutanese *S. sonnei* MDR by 2006. Furthermore, a fluoroquinolone-resistant subclade within the Central Asia clade was predicted to have arisen between 1999 and 2006. In addition, the fixation of a novel recombinant colicin E plasmid (pSSE3) may have offered this group of pathogens a competitive advantage over other enteric bacteria, similar to the colicin-associated selective sweep observed in *S. sonnei* in Vietnam (Holt *et al.*, 2013). These data suggest that whilst antimicrobial resistance is clearly important, it is not sufficient in itself for driving the spread of new clones and anticompetitive functions appear to be key themes in the spread of MDR *S. sonnei*.

Our data additionally suggests that international travel may have resulted in the dissemination of Central Asia clade *S. sonnei* beyond Bhutan and the remainder of the Indian subcontinent. *S. sonnei* isolates with similar PFGE patterns to the strains in this analysis have been recovered from cases of travel-associated shigellosis in Ireland (De Lappe *et al.*, 2015) and in a *Shigella* outbreak amongst 'men who have sex with men' in Canada (Gaudreau *et al.*, 2011). The *S. sonnei* isolated from both instances were found to be ciprofloxacin-resistant. Studies conducted in Japan have established a strong association between this pulsotype and travel to India, and increasing prevalence of nalidixic acid-resistant *S. sonnei* has been observed since 2001 (Uchimura *et al.*, 2001; Izumiya *et al.*, 2009). Shigellosis surveys in Iran conducted from 2008 to 2012 have repeatedly shown that this characteristic Asian pulsotype is prevalent and endemic, and that these strains also possess a similar class 2 integron (*dfrA1*, *sat1*), but exhibit variability in their ciprofloxacin susceptibility phenotype (Tajbakhsh *et al.*, 2012; Eftekhari *et al.*, 2013; Alizadeh-Hesar *et al.*, 2015). The first case of fluoroquinolone-resistant *S. sonnei* was identified in Japan in 1993 (Horiuchi *et al.*, 1993). Since then, the incidence of ciprofloxacin-resistant *Shigella* in Asia and Africa has risen dramatically, from 3.9 % in 2001–2003 to 29.1 % in 2007–2009 (Gu *et al.*, 2012). The potential emergence and dissemination of fluoroquinolone-resistant *S. sonnei* poses a substantial threat to public health, including low shigellosis burden areas in developed nations, as again substantiated by the recent fluoroquinolone-resistant *S. sonnei* outbreak in the USA (Bowen *et al.*, 2015). Taken together, these data strongly suggest that the Central Asia clade of *S. sonnei* is widely present across South Asia and beyond, and may represent an emerging threat to public health.

Acknowledgements

We wish to acknowledge the staff of the Department of Enteric Disease, AFRIMS, Bangkok, Thailand for their work in isolating and maintaining the strains in this study. S. B. is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and Royal Society (100087/Z/12/Z). N. R. T. is supported the Wellcome Trust grant 098051 to Wellcome Trust Sanger Institute. The views expressed in this article are those of the authors and do not represent the official policy or position of the US Department of the Army, Department of Defense or US Government.

References

Alizadeh-Hesar, M., Bakhshi, B. & Najar-Peerayeh, S. (2015). Clonal dissemination of a single *Shigella sonnei* strain among Iranian children during Fall 2012 in Tehran, I.R. Iran. *Infect Genet Evol* **34**, 260–266.

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M. A. & Alekseyenko, A. V. (2012). Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol* **29**, 2157–2167.

Baele, G., Li, W. L., Drummond, A. J., Suchard, M. A. & Lemey, P. (2013). Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Mol Biol Evol* **30**, 239–243.

Baker, K. S., Dallman, T. J., Ashton, P. M., Day, M., Hughes, G., Crook, P. D., Gilbart, V. L., Zittermann, S., Allen, V. G. & other authors (2015). Intercontinental dissemination of azithromycin-resistant shigellosis through sexual transmission: a cross-sectional study. *Lancet Infect Dis* **15**, 913–921.

Berry, J.-L. & Pelicic, V. (2015). Exceptionally widespread nanomachines composed of type IV pilins: the prokaryotic Swiss Army knives. *FEMS Microbiol Rev* **39**, 134–154.

Ongpin, M. & other authors (2015). Importation and domestic transmission of *Shigella sonnei* resistant to ciprofloxacin – United States, May 2014–February 2015. *MMWR Morb Mortal Wkly Rep* **64**, 318–320.

Calcuttawala, F., Hariharan, C., Pazhani, G. P., Ghosh, S. & Ramamurthy, T. (2015). Activity spectrum of colicins produced by *Shigella sonnei* and genetic mechanism of colicin resistance in conspecific *S. sonnei* strains and *Escherichia coli*. *Antimicrob Agents Chemother* **59**, 152–158.

Carver, T., Berriman, M., Tivey, A., Patel, C., Böhme, U., Barrell, B. G., Parkhill, J. & Rajandream, M. A. (2008). Artemis and act: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* **24**, 2672–2676.

CLSI (2012). *Performance Standards for Antimicrobial Susceptibility Testing*; 20th Informational Supplement. Wayne, PA: Clinical and Laboratory Standards Institute.

Connor, T. R., Barker, C. R., Baker, K. S., Weill, F. X., Talukder, K. A., Smith, A. M., Baker, S., Gouali, M., Pham Thanh, D. & other authors (2015). Species-wide whole genome sequencing reveals historical global spread and recent local persistence in *Shigella flexneri*. *eLife* **4**, 1–16.

Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., Parkhill, J. & Harris, S. R. (2014). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* **43**, e15.

Chung The, H., Karkey, A., Pham Thanh, D., Boinett, C. J., Cain, A. K., Ellington, M., Baker, K. S., Dongol, S., Thompson, C. & other authors (2015). A high-resolution genomic analysis of multidrug-resistant hospital outbreaks of *Klebsiella pneumoniae*. *EMBO Mol Med* **7**, 227–239.

De Lappe, N., O'Connor, J., Garvey, P., McKeown, P. & Cormican, M. (2015). Ciprofloxacin-resistant *Shigella sonnei* associated with travel to India. *Emerg Infect Dis* **21**, 894–896.

Downloaded from www.microbiologyresearch.org by
IP: 118.69.34.244
On: Mon, 28 Dec 2015 07:55:31

Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214.

Dudley, E. G., Abe, C., Ghigo, J. M., Latour-Lambert, P., Hormazabal, J. C. & Nataro, J. P. (2006). An IncI1 plasmid contributes to the adherence of the atypical enteroaggregative *Escherichia coli* strain C1096 to cultured cells and abiotic surfaces. *Infect Immun* **74**, 2102–2114.

Eftekhari, N., Bakhshi, B., Pourshafie, M. R., Zarbakhsh, B., Rahbar, M., Hajia, M. & Ghazvini, K. (2013). Genetic diversity of *Shigella* spp. and their integron content. *Foodborne Pathog Dis* **10**, 237–242.

Gaudreau, C., Ratnayake, R., Pilon, P. A., Gagnon, S., Roger, M. & Lévesque, S. (2011). Ciprofloxacin-resistant *Shigella sonnei* among men who have sex with men, Canada, 2010. *Emerg Infect Dis* **17**, 1747–1750.

Gu, B., Cao, Y., Pan, S., Zhuang, L., Yu, R., Peng, Z., Qian, H., Wei, Y., Zhao, L. & other authors (2012). Comparison of the prevalence and changing resistance to nalidixic acid and ciprofloxacin of Shigella between Europe-America and Asia-Africa from 1998 to 2009. *Int J Antimicrob Agents* **40**, 9–17.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307–321.

Han, J., Lynne, A. M., David, D. E., Tang, H., Xu, J., Nayak, R., Kaldhone, P., Logue, C. M. & Foley, S. L. (2012). DNA sequence analysis of plasmids from multidrug resistant *Salmonella enterica* serotype Heidelberg isolates. *PLoS One* **7**, e51160.

Holt, K. E., Baker, S., Weill, F. X., Holmes, E. C., Kitchen, A., Yu, J., Sangal, V., Brown, D. J., Coia, J. E. & other authors (2012). *Shigella sonnei* genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. *Nat Genet* **44**, 1056–1059.

Holt, K. E., Thieu Nga, T. V., Thanh, D. P., Vinh, H., Kim, D. W., Vu Tra, M. P., Campbell, J. I., Hoang, N. V., Vinh, N. T. & other authors (2013). Tracking the establishment of local endemic populations of an emergent enteric pathogen. *Proc Natl Acad Sci U S A* **110**, 17522–17527.

Horiuchi, S., Inagaki, Y., Yamamoto, N., Okamura, N., Imagawa, Y. & Nakaya, R. (1993). Reduced susceptibilities of *Shigella sonnei* strains isolated from patients with dysentery to fluoroquinolones. *Antimicrob Agents Chemother* **37**, 2486–2489.

Huang, I. F., Chiu, C. H., Wang, M. H., Wu, C. Y., Hsieh, K. S. & Chiou, C. C. (2005). Outbreak of dysentery associated with ceftriaxone-resistant *Shigella sonnei*: first report of plasmid-mediated CMY-2-type AmpC beta-lactamase resistance in *S. sonnei*. *J Clin Microbiol* **43**, 2608–2612.

Izumiya, H., Tada, Y., Ito, K., Morita-Ishihara, T., Ohnishi, M., Terajima, J. & Watanabe, H. (2009). Characterization of *Shigella sonnei* isolates from travel-associated cases in Japan. *J Med Microbiol* **58**, 1486–1491.

Kentner, D., Martano, G., Callon, M., Chiquet, P., Brodmann, M., Burton, O., Wahlander, A., Nanni, P., Delmotte, N. & other authors (2014). *Shigella* reroutes host cell central metabolism to obtain high-flux nutrient supply for vigorous intracellular growth. *Proc Natl Acad Sci U S A* **111**, 9929–9934.

Kim, S. R. & Komano, T. (1997). The plasmid R64 thin pilus identified as a type IV pilus. *J Bacteriol* **179**, 3594–3603.

Kleanthous, C. (2010). Swimming against the tide: progress and challenges in our understanding of colicin translocation. *Nat Rev Microbiol* **8**, 843–848.

Kotloff, K. L., Blackwelder, W. C., Nasrin, D., Nataro, J. P., Farag, T. H., van Eijk, A., Adegbola, R. A., Alonso, P. L., Breiman, R. F. & other authors (2012). The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiologic and clinical methods of the case/control study. *Clin Infect Dis* **55** (Suppl 4), S232–S245.

Kotloff, K. L., Nataro, J. P., Blackwelder, W. C., Nasrin, D., Farag, T. H., Panchalingam, S., Wu, Y., Sow, S. O., Sur, D. & other authors (2013). Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* **382**, 209–222.

Künne, C., Billion, A., Mshana, S. E., Schmiedel, J., Domann, E., Hossain, H., Hain, T., Imirzalioglu, C. & Chakraborty, T. (2012). Complete sequences of plasmids from the hemolytic-uremic syndrome-associated Escherichia coli strain HUSEC41. *J Bacteriol* **194**, 532–533.

Levine, M. M., Kotloff, K. L., Barry, E. M., Pasetti, M. F. & Sztein, M. B. (2007). Clinical trials of *Shigella* vaccines: two steps forward and one step back on a long, hard road. *Nat Rev Microbiol* **5**, 540–553.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009). 1000 Genome Project Data Processing Subgroup The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.

Liu, L., Johnson, H. L., Cousens, S., Perin, J., Scott, S., J.E., Rudan, I., Campbell, H., Cibulskis, R. & other authors (2012). Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet* **379**, 2151–2161.

Marfin, A. A., Moore, J., Collins, C., Biellik, R., Kattel, U., Toole, M. J. & Moore, P. S. (1994). Infectious disease surveillance during emergency relief to Bhutanese refugees in Nepal. *JAMA* **272**, 377–381.

Nguyen, N. T. K., Ha, V., Tran, N. V., Stabler, R., Pham, D. T., Le, T. M., van Doorn, H. R., Cerdeño-Tárraga, A., Thomson, N. & other authors (2010). The sudden dominance of bla_{CTX-M} harbouring plasmids in *Shigella* spp., Circulating in Southern Vietnam. *PLoS Negl Trop Dis* **4**, e702.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol Biol Evol* **25**, 1253–1256.

Prosseda, G., Di Martino, M. L., Campilongo, R., Fioravanti, R., Micheli, G., Casalino, M. & Colonna, B. (2012). Shedding of genes that interfere with the pathogenic lifestyle: the *Shigella* model. *Res Microbiol* **163**, 399–406.

Pu, X. Y., Pan, J. C., Zhang, W., Zheng, W., Wang, H. Q. & Gu, Y. M. (2015). Quinolone resistance-determining region mutations and the plasmid-mediated quinolone resistance gene *qnrS* played important roles in decreased susceptibility to fluoroquinolones among *Shigella* isolates in southeast China between 1998 and 2013. *Int J Antimicrob Agents* **45**, 438–439.

Pupo, G. M., Lan, R. & Reeves, P. R. (2000). Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc Natl Acad Sci U S A* **97**, 10567–10572.

Quail, M. A., Kozarewa, I., Smith, F., Scally, A., Stephens, P. J., Durbin, R., Swerdlow, H. & Turner, D. J. (2008). A large genome center's improvements to the Illumina sequencing system. *Nat Methods* **5**, 1005–1010.

Rohmer, L., Jacobs, M. A., Brittnacher, M. J., Fong, C., Hayden, H. S., Hocquet, D., Weiss, E. J., Radey, M., Germani, Y. & other authors (2014). Genomic analysis of the emergence of 20th century epidemic dysentery. *BMC Genomics* **15**, 355.

Ruekit, S., Wangchuk, S., Dorji, T., Tshering, K. P., Pootong, P., Nobthai, P., Serichantalergs, O., Poramathikul, K., Bodhidatta, L. & Mason, C. J. (2014). Molecular characterization and PCR-based replicon typing of multidrug resistant *Shigella sonnei* isolates from an outbreak in Thimphu, Bhutan. *BMC Res Notes* **7**, 95.

Sansonetti, P. J., Kopecko, D. J. & Formal, S. B. (1982). Involvement of a plasmid in the invasive ability of *Shigella flexneri*. *Infect Immun* **35**, 852–860.

Downloaded from www.microbiologyresearch.org by
IP: 118.69.34.244
On: Mon, 28 Dec 2015 07:55:31

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069.

Seol, S. Y., Kim, Y. T., Jeong, Y. S., Oh, J. Y., Kang, H. Y., Moon, D. C., Kim, J., Lee, Y. C., Cho, D. T. & Lee, J. C. (2006). Molecular characterization of antimicrobial resistance in *Shigella sonnei* isolates in Korea. *J Med Microbiol* **55**, 871–877.

Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G. & Ng, P. C. (2012). SIFT WEB SERVER: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* **40** (W1), W452–W457.

Tajbakhsh, M., García Migura, L., Rahbar, M., Svendsen, C. A., Mohammadzadeh, M., Zali, M. R., Aarestrup, F. M. & Hendriksen, R. S. (2012). Antimicrobial-resistant *Shigella* infections from Iran: an overlooked problem? *J Antimicrob Chemother* **67**, 1128–1133.

Thompson, C. N., Duy, P. T. & Baker, S. (2015). The rising dominance of *Shigella sonnei*: an intercontinental shift in the etiology of bacillary dysentery. *PLoS Negl Trop Dis* **9**, e0003708.

Uchimura, M., Kishida, K. & Koiwai, K. (2001). [Increasing incidence and the mechanism of resistance of nalidixic acid resistant *Shigella sonnei*]. *Kansenshogaku Zasshi* **75**, 923–930, (in Japanese).

Vinh, H., Nhu, N. T., Nga, T. V., Duy, P. T., Campbell, J. I., Hoang, N. V., Boni, M. F., My, P. V., Parry, C. & other authors (2009). A changing picture of shigellosis in southern Vietnam: shifting species dominance, antimicrobial susceptibility and clinical presentation. *BMC Infect Dis* **9**, 204.

Walker, C. L., Rudan, I., Liu, L., Nair, H., Theodoratou, E., Bhutta, Z. A., O'Brien, K. L., Campbell, H. & Black, R. E. (2013). Global burden of childhood pneumonia and diarrhoea. *Lancet* **381**, 1405–1416.

WHO (2014). World Health Statistics. http://apps.who.int/iris/bitstream/10665/112738/1/9789240692671_eng.pdf

Zerbino, D. R. & Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* **18**, 821–829.

Zwickl, D. J. (2006). *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion.* PhD thesis, University of Texas at Austin, Austin, TX, USA.

Data Bibliography

1. *Shigella sonnei* Ss046 complete genome: GenBank accession number NC_007384.

2. *Shigella sonnei* strain IDH01791 plasmid pSSE3 complete sequence: GenBank accession number KP970685.

3. *Escherichia coli* plasmid pHUSEC41-1 complete sequence: GenBank accession number NC_018995.

4. *Escherichia coli* HUSEC2011 plasmid pHUSEC2011-1 complete sequence: GenBank accession number NC_022742.

5. *Escherichia coli* plasmid pSERB1 partial sequence: GenBank accession number NG_035985.

6. *Salmonella enterica* subsp. enterica serovar Heidelberg plasmid pSH146_65 complete sequence: GenBank accession number NC_019115.

7. Table S1. http://figshare.com/articles/The_introduction_and_establishment_of_fluoroquinolone_resistant_*Shigella*_sonnei_into_Bhutan/1610693

Downloaded from www.microbiologyresearch.org by
IP: 118.69.34.244
On: Mon, 28 Dec 2015 07:55:31

PLOS | MEDICINE

RESEARCH ARTICLE

South Asia as a Reservoir for the Global Spread of Ciprofloxacin-Resistant *Shigella sonnei*: A Cross-Sectional Study

Hao Chung The[1], Maia A. Rabaa[1,2], Duy Pham Thanh[1], Niall De Lappe[3], Martin Cormican[4], Mary Valcanis[5], Benjamin P. Howden[5], Sonam Wangchuk[6], Ladaporn Bodhidatta[7], Carl J. Mason[7], To Nguyen Thi Nguyen[1], Duong Vu Thuy[1], Corinne N. Thompson[1,2], Nguyen Phu Huong Lan[1,8], Phat Voong Vinh[1], Tuyen Ha Thanh[1], Paul Turner[2,9], Poda Sar[9], Guy Thwaites[1,2], Nicholas R. Thomson[10,11], Kathryn E. Holt[12,13‡], Stephen Baker[1,2,11‡*]

1 The Hospital for Tropical Diseases, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam, **2** Centre for Tropical Medicine and Global Health, Oxford University, Oxford, United Kingdom, **3** National *Salmonella*, *Shigella*, and *Listeria monocytogenes* Reference Laboratory, University Hospital Galway, Galway, Ireland, **4** School of Medicine, National University of Ireland Galway, Galway, Ireland, **5** Microbiological Diagnostic Unit Public Health Laboratory, Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, The University of Melbourne, Melbourne, Australia, **6** Public Health Laboratory, Department of Public Health, Ministry of Health, Royal Government of Bhutan, Thimphu, Bhutan, **7** Department of Enteric Diseases, Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand, **8** The Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam, **9** Cambodia-Oxford Medical Research Unit, Angkor Hospital for Children, Siem Reap, Cambodia, **10** The London School of Hygiene and Tropical Medicine, London, United Kingdom, **11** The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom, **12** Centre for Systems Genomics, The University of Melbourne, Melbourne, Australia, **13** Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Melbourne, Australia

‡ These authors are joint senior authors on this work.
* sbaker@oucru.org

CrossMark
click for updates

OPEN ACCESS

Citation: Chung The H, Rabaa MA, Pham Thanh D, De Lappe N, Cormican M, Valcanis M, et al. (2016) South Asia as a Reservoir for the Global Spread of Ciprofloxacin-Resistant *Shigella sonnei*: A Cross-Sectional Study. PLoS Med 13(8): e1002055. doi:10.1371/journal.pmed.1002055

Academic Editor: Lorenz von Seidlein, Mahidol-Oxford Tropical Medicine Research Unit, THAILAND

Received: October 13, 2015

Accepted: May 18, 2016

Published: August 2, 2016

Copyright: © 2016 Chung The et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Sequences available via the GenomeTrackr Project (http://www.ncbi.nlm.nih.gov/sra/; project number: PRJNA218110). All sequences were mapped to the S. sonnei Ss046 reference sequence (http://www.ncbi.nlm.nih.gov/nuccore/; accession number: NC_007384). Global analysis accession: ERP00018. The related metadata for the bacterial isolates are shown in S1 Table.

Funding: SB is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (100087/Z/12/Z). NRT is supported by the Wellcome Trust grant #098051 to Wellcome Trust Sanger

Abstract

Background

Antimicrobial resistance is a major issue in the *Shigellae*, particularly as a specific multi-drug-resistant (MDR) lineage of *Shigella sonnei* (lineage III) is becoming globally dominant. Ciprofloxacin is a recommended treatment for *Shigella* infections. However, ciprofloxacin-resistant *S. sonnei* are being increasingly isolated in Asia and sporadically reported on other continents. We hypothesized that Asia is a primary hub for the recent international spread of ciprofloxacin-resistant *S. sonnei*.

Methods and Findings

We performed whole-genome sequencing on a collection of 60 contemporaneous ciprofloxacin-resistant *S. sonnei* isolated in four countries within Asia (Vietnam, $n = 11$; Bhutan, $n = 12$; Thailand, $n = 1$; Cambodia, $n = 1$) and two outside of Asia (Australia, $n = 19$; Ireland, $n = 16$). We reconstructed the recent evolutionary history of these organisms and combined these data with their geographical location of isolation. Placing these sequences into

Institute. KEH is supported by fellowship #1061409 from the NHMRC of Australia. BPH is supported by fellowship #1023526 from the NHMRC of Australia. 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.

Abbreviations: AHC, Angkor Hospital for Children; GEMS, Global Enteric Multicentre Study; MDR, multidrug-resistant; MIC, minimum inhibitory concentration; MRE, extended majority rule; OxTREC, Oxford Tropical Research Ethics Committee; PFGE, pulsed field gel electrophoresis; QRDR, quinolone resistance-determining region; REBH, Research Ethics Board of Health; WRAIR, Water Reed Army Institute of Research.

a global phylogeny, we found that all ciprofloxacin-resistant *S. sonnei* formed a single clade within a Central Asian expansion of lineage III. Furthermore, our data show that resistance to ciprofloxacin within *S. sonnei* may be globally attributed to a single clonal emergence event, encompassing sequential *gyrA*-S83L, *parC*-S80I, and *gyrA*-D87G mutations. Geographical data predict that South Asia is the likely primary source of these organisms, which are being regularly exported across Asia and intercontinentally into Australia, the United States and Europe. Our analysis was limited by the number of *S. sonnei* sequences available from diverse geographical areas and time periods, and we cannot discount the potential existence of other unsampled reservoir populations of antimicrobial-resistant *S. sonnei*.

Conclusions

This study suggests that a single clone, which is widespread in South Asia, is likely driving the current intercontinental surge of ciprofloxacin-resistant *S. sonnei* and is capable of establishing endemic transmission in new locations. Despite being limited in geographical scope, our work has major implications for understanding the international transfer of antimicrobial-resistant pathogens, with *S. sonnei* acting as a tractable model for studying how antimicrobial-resistant Gram-negative bacteria spread globally.

Author Summary

Why Was This Study Done?

- Antimicrobial resistance is a major issue in *Shigella*, and ciprofloxacin is a recommended treatment for *Shigella* infections.
- Ciprofloxacin-resistant *Shigella sonnei* are being increasingly isolated globally.

What Did the Researchers Do and Find?

- We performed genome sequencing on 60 ciprofloxacin-resistant *S. sonnei* isolated in six countries within and outside of Asia.
- Placing these genome sequences in context with other strains, we found that all ciprofloxacin-resistant *S. sonnei* formed a single clade, with South Asia as the likely primary source of these organisms.
- These organisms are also being regularly exported across Asia and intercontinentally into Australia, the United States, and Europe.
- The number of *S. sonnei* sequences available from other locations limited our analysis.

What Do These Findings Mean?

- Our work suggests that a single clone, which is widespread in South Asia, is driving the current global surge of ciprofloxacin-resistant *S. sonnei*.
- We argue that *S. sonnei* acts as a tractable model for studying how antimicrobial-resistant, Gram-negative bacteria spread globally.

Introduction

Diarrheal disease is the second most common cause of mortality in children under the age of 5 y worldwide, equating to approximately 800,000 deaths per year [1]. The recent Global Enteric Multicentre Study (GEMS), a large, prospective, case-control study focusing on mild and severe paediatric diarrheal illnesses in sub-Saharan Africa and South Asia, found that *Shigella* (a genus of Gram-negative enteric bacteria) were amongst the top four most prevalent diarrhoeal pathogens in these settings [2]. The most recent estimates suggest that *Shigella* infections account for around 125 million cases of diarrhoea annually, with the majority occurring in children in low-income countries [3]. There are four *Shigella* species (*dysenteriae*, *boydii*, *flexneri*, and *sonnei*), but the overwhelming majority of the current global burden is presently caused by *S. sonnei* and *S. flexneri*. Present-day international epidemiology of the various *Shigella* species is particularly intriguing, as *S. sonnei* is replacing *S. flexneri* as the most common cause of shigellosis worldwide; this pattern is accentuated in regions undergoing rapid economic development [4,5], where *S. flexneri* dominated as recently as a decade ago.

Shigella infections are characterised by the invasion and disruption of the epithelial cells lining the gastrointestinal mucosa, resulting in mucous and/or bloody diarrhoeal discharge. Although shigellosis is typically self-limiting, antimicrobial treatment is used to prevent complications, reduce dysenteric discharge, and curb post-symptomatic faecal shedding [6,7]. Consequently, resistance to antimicrobials restricts treatment options, placing vulnerable individuals suffering from shigellosis at increased risk of complications and increasing the likelihood of protracted faecal shedding. One of the current recommended first-line treatments for shigellosis is the fluoroquinolone ciprofloxacin [8]. The fluoroquinolones target the DNA gyrase, a type II topoisomerase that is essential for bacterial DNA replication and transcription [9].

Antimicrobial resistance is an emerging global issue in *S. sonnei*, with a specific multidrug-resistant (MDR) lineage (III) now dominating internationally. Furthermore, organisms belonging to lineage III appear to be highly proficient at acquiring resistance to additional antimicrobials (including third-generation cephalosporins) when they are introduced into new locations [10]. However, given their common usage and broad spectrum of activity, resistance against the fluoroquinolones is the most concerning. Since the first isolation of *S. sonnei* with reduced susceptibility to ciprofloxacin in Japan in 1993 [11], ciprofloxacin-resistant *S. sonnei* have been increasingly reported throughout Asia [12–14]. Furthermore, public health laboratories in several non-Asian countries with low incidences of shigellosis have reported the isolation of ciprofloxacin-resistant *S. sonnei*, often from individuals reporting recent travel to locations with a high risk of shigellosis [15–17].

Whole-genome sequencing has proven to be the gold standard for tracking the international dissemination of clonal bacterial pathogens [18,19], and we have previously exploited

this method to study the phylogenetic structure and spread of *S. sonnei* at both national and intercontinental levels [10,20]. Hypothesising that Asia was a hub for the recent international spread of ciprofloxacin-resistant *S. sonnei*, we performed whole-genome sequencing and phylogenetic characterisation of a collection of ciprofloxacin-resistant *S. sonnei* isolated from within and outside Asia, aiming to explore the origins of this growing international epidemic.

Methods

Ethics Statement

S. sonnei isolates from Bhutan, Thailand, and Vietnam were collected during diarrheal surveillance studies [14]. IRB approval was granted for these studies (including organism characterization) from the Research Ethics Board of Health (REBH), Ministry of Health, Bhutan (Bhutan study), the Walter Reed Army Institute of Research (WRAIR) Institutional Review Board, USA (Bhutan and Thailand studies), the Oxford Tropical Research Ethics Committee (OxTREC), UK, and the Hospital for Tropical Diseases Ho Chi Minh City, Vietnam (Vietnam study). Written (Vietnam) or oral (Thailand and Bhutan) consent was obtained from a parent or guardian at the time of enrolment into the study. The target patient groups for these studies were generally hospitalised children aged less than 5 y residing in close proximity to the study centres. *S. sonnei* isolates from Cambodia were collected at the Angkor Hospital for Children (AHC) in Siem Reap province from the routine diagnostic laboratory; no patient data were collected, and these organisms were not subject to local or international IRB approval. The ciprofloxacin-resistant *S. sonnei* from countries outside Asia were collected and characterized by the National *Salmonella*, *Shigella*, and *Listeria monocytogenes* Reference Laboratory, Galway, Ireland, and the Microbiological Diagnostic Unit Public Health Laboratory, Melbourne, Australia. These isolates were generally, but not exclusively, obtained from patients reporting recent travel to countries with a high incidence of shigellosis in Asia (Table 1). As these isolates were from anonymous sources and collected at local public health laboratories, these were not subject to IRB approval, and informed consent was not required.

Strain Collection

Aiming to investigate the current international upsurge in ciprofloxacin-resistant *S. sonnei* in detail, we gathered a collection of 60 contemporary ciprofloxacin-resistant *S. sonnei* from six countries for whole-genome sequencing. The isolates originated from Asian countries with a high incidence of shigellosis (Vietnam, $n = 11$; Bhutan, $n = 12$; Thailand, $n = 1$; Cambodia, $n = 1$), as well as countries with a low incidence of shigellosis (Australia, $n = 19$; Ireland, $n = 16$). Twelve additional ciprofloxacin-susceptible *S. sonnei* sequences from these settings were also included for phylogenetic context. All strains were isolated independently between 2010 and 2015; details of the isolates used in this study are shown in Table 1.

Susceptibility to ciprofloxacin was determined by disk diffusion, E-test, agar dilution, or broth microdilution, depending on the collaborating institution, and susceptibility breakpoints were interpreted according to the European Committee on Antimicrobial Susceptibility Testing (http://www.eucast.org/clinical_breakpoints). Namely, resistance was determined as strains with a zone of inhibition <19 mm (5 μg disc) and/or a minimum inhibitory concentration (MIC) >1 μg/ml against ciprofloxacin; the various location-specific methods and resulting data are described in Table 1.

Table 1. The origins of the *Shigella sonnei* isolates and sequences used in this study.

Country	Isolates in Central Asia clade (N)	Ciprofloxacin-resistant isolates (N)	Study or institute origin	IRB approval or public database access	Patient group	Region of recent travel history (N)	Sequencing platform/ Public database	Ciprofloxacin susceptibility
Bhutan	12	12	Diarrhoeal disease surveillance in JDWNRH[a], Thimphu, Bhutan (AFRIMS[b])	The Research Ethics Board of Health (REBH), Ministry of Health, Bhutan, and the Walter Reed Army Institute of Research (WRAIR) Institutional Review Board, USA	Hospitalised children <5 y old	NA	Illumina HiSeq 2000	Disk diffusion/E-test
Vietnam	11	11	Diarrhoeal disease surveillance in Ho Chi Minh City, Vietnam (OUCRU[c])	Oxford Tropical Research Ethics Committee (OxTREC), UK, and the Hospital for Tropical Diseases Ho Chi Minh City, Vietnam	Hospitalised children <5 y old	NA	Illumina MiSeq	Disk diffusion
Thailand	8	1	Diarrhoeal disease surveillance in Thailand (AFRIMS)	The Research Ethics Board of Health (REBH), Ministry of Health, Bhutan, and the Walter Reed Army Institute of Research (WRAIR) Institutional Review Board, USA	Hospitalised children <5 y old	NA	Illumina HiSeq 2000	Disk diffusion
Cambodia	1	1	Diarrhoeal disease surveillance in Siem Reap, Cambodia (COMRU[d])	Anonymous clinical isolates collected as part of routine diagnosis—IRB approval not required	Hospitalised children <5 y old	NA	Illumina HiSeq 2000	Disk diffusion
Ireland	20	16	National *Salmonella*, *Shigella*, and *L. monocytogenes* Reference Laboratory, Galway, Ireland	Anonymous clinical isolates sent to public health reference laboratory—IRB approval not required	Primarily patients with recent travel history	India (9), Germany (1), Morocco (1), No travel (5), Unknown (4)	Illumina HiSeq 2000	Broth microdilution
Australia	20	19	Microbiological Diagnostic Unit Public Health Laboratory in Melbourne, Australia	Anonymous clinical isolates sent to public health reference laboratory—IRB approval not required	Patients with recent travel history	India (15), Cambodia (3), Thailand (1), Southeast Asia (1).	Illumina NextSeq	Agar dilution
United States	14	10	Genome Trackr; Centre for Disease Control and Prevention, USA	Data accessed from NCBI under the BioProject accession number PRJNA218110	Unknown	Unknown	Illumina MiSeq	in silico assessment of QRDR[e] mutations

[a] Jigme Dorji Wangchuk National Referral Hospital

[b] Armed Forces Research Institute of Medical Sciences

[c] Oxford University Clinical Research Unit

[d] Cambodia Oxford Medical Research Unit

[e] quinolone resistance-determining region

NA, Not applicable

doi:10.1371/journal.pmed.1002055.t001

Genome Sequencing and Analysis

All isolated *S. sonnei* were subcultured and subjected to DNA extraction prior to whole-genome sequencing on various Illumina platforms to produce pair-ended short read sequences; the specific sequencing system and the resulting public database numbers are shown in Table 1. We additionally included 14 *S. sonnei* sequences from the NCBI database under the Bioproject accession number PRJNA218110. These organisms were isolated in the US and deposited as part of the GenomeTrackr Project. All sequences were mapped to the *S. sonnei* Ss046 reference sequence (Accession number: NC_007384) using SMALT (version 0.7.4), and SNPs were called against the reference and filtered using SAMtools [21]. To contextualize all ciprofloxacin-resistant *S. sonnei* within the global phylogeny, we appended our collection to include 133 publicly available sequences from a previous global analysis (accession ERP000182) [20]. Previously characterized mobile genetic elements and putative recombination (predicted using Gubbins) were removed [20], resulting in a gap-free alignment of 211 non-duplicate pseudo-whole genome sequences of 4,738 SNPs. A whole-genome phylogeny was inferred from this alignment using RAxML v8.1.3 under the GTRGAMMA substitution model, and sufficient bootstrap replicates were determined automatically using the extended majority rule (MRE) bootstrap convergence criterion. Bootstrap values >75% signify strong statistical support for a node within a maximum likelihood phylogeny, thus indicating that all isolates falling within that lineage are highly likely to be linked in evolutionary history at a more recent time than those falling outside of that lineage. In order to obtain a refined phylogenetic structure of the Central Asia clade, we applied the aforementioned approach to a set of 97 *S. sonnei* sequences (86 novel sequences and 11 historical sequences) belonging to this clade. This resulted in an alignment of 1,121 SNPs, which was used for phylogenetic inference. De novo assemblies were generated for each read set using Velvet and VelvetOptimiser, and read sets were mapped back to each assembly [22]. A manually curated database based on ResFinder [23] was mapped against each assembly to identify mobile resistance genetic determinants in all ciprofloxacin-resistant strains.

Results

Fluoroquinolone-Resistant *S. sonnei* in a Global Context

We constructed a whole genome phylogeny of *S. sonnei*, incorporating sequences from 133 globally representative isolates and 86 novel isolates from Vietnam, Cambodia, Thailand, Bhutan, Australia, Ireland, and the US. The novel sequences included 60 from ciprofloxacin-resistant (MIC >1 μg/ml or zone of inhibition <19 mm) organisms and 26 from ciprofloxacin-susceptible organisms (or of unknown ciprofloxacin susceptibility isolated in the US). The related metadata for the bacterial isolates are shown in S1 Table. The overall tree topology reflected the previously described global phylogenetic structure [20], confirming the presence of four distinct lineages (I, II, III, and IV); lineage III was the most commonly represented and the most widely geographically distributed (Fig 1A). All ciprofloxacin-resistant *S. sonnei* formed a single, well-supported monophyletic clade within the Central Asian expansion of lineage III (Central Asia III), an MDR group that is closely related but distinct from the Global III clade (Fig 1A and 1B).

The Emergence of a Fluoroquinolone-Resistant *S. sonnei* Clone

We next performed a more detailed phylogenetic reconstruction of the Central Asia III clade, incorporating sequence data from the 60 phenotypically ciprofloxacin-resistant isolates and 26 others (ciprofloxacin-susceptible or of unknown ciprofloxacin susceptibility), along with 11

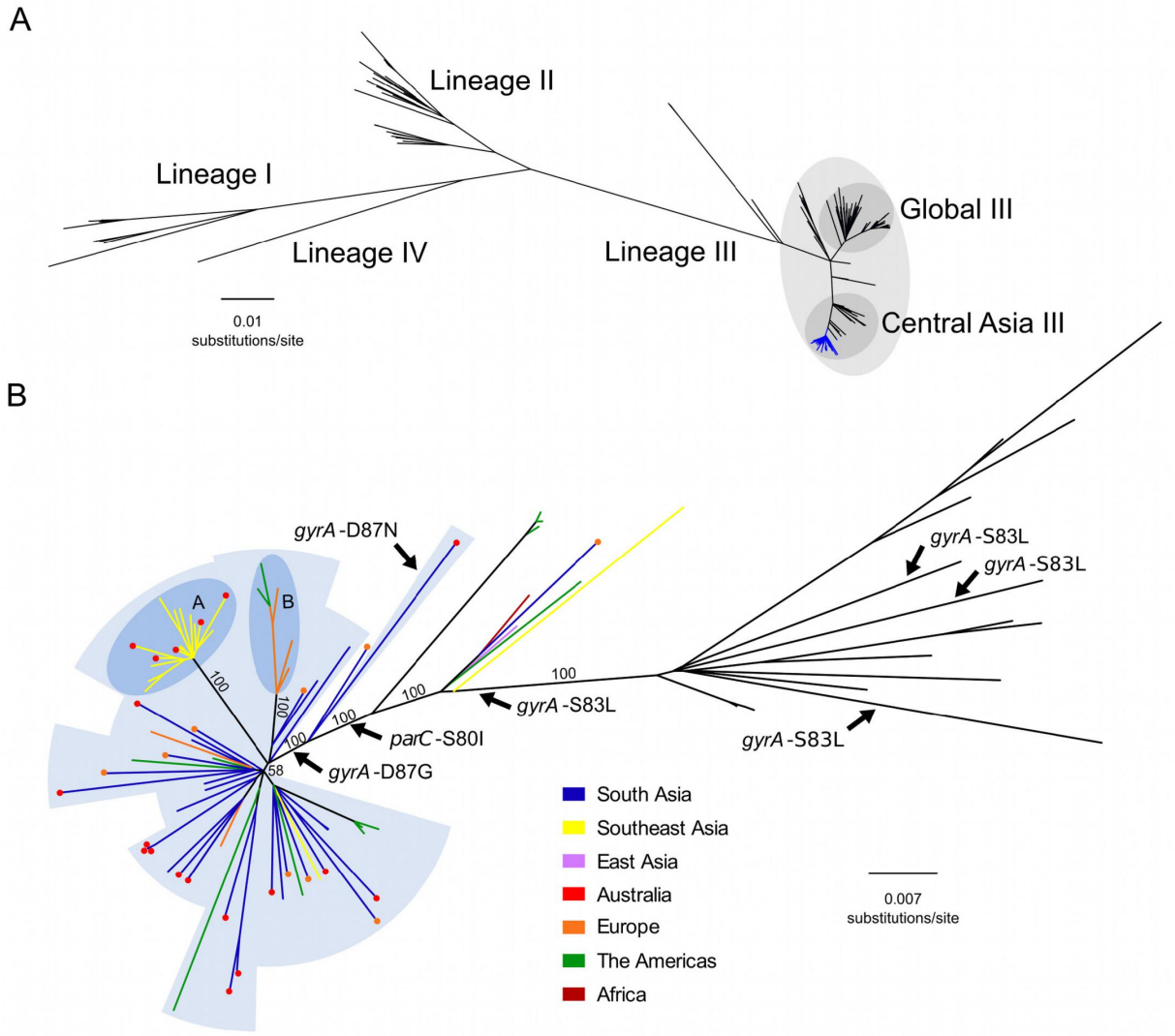


Fig 1. The phylogenetic structure of ciprofloxacin-resistant *Shigella sonnei* in an international context. A) Unrooted maximum likelihood phylogeny of 211 globally representative *S. sonnei*, including 60 sequences from ciprofloxacin-resistant isolates (highlighted by the blue branches). Major lineages are indicated by numbers (I, II, III, and IV) as defined in Holt et al. 2012, with clades Global III and Central Asia III within lineage III highlighted. Horizontal scale bar indicates the number of nucleotide substitutions per site. B) Unrooted maximum likelihood phylogeny of Central Asia III, composed of 97 *S. sonnei* sequences. Branch colours indicate region of isolation (where no travel history is confirmed) or region of recent travel (where travel history was confirmed) according to the keys. For isolates with confirmed recent travel, a coloured circle at the tip indicates the region where the isolate was collected (multiple coloured circles are indicative of multiple isolates). Labelled arrows indicate branches where the mutations *gyrA*-S83L, *gyrA*-D87N, *gyrA*-D87G, and *parC*-S80I have arisen. Blue background shading denotes isolates exhibiting ciprofloxacin resistance conferred by triple mutations (*gyrA*-S83L, *parC*-S80I, and *gyrA*-D87G [or *gyrA*-D87N]). Subpopulations A and B are highlighted in the darker blue shaded areas, denoting clonal expansions in Southeast Asia and Europe/America, respectively. Numbers above major branches indicate bootstrap support values, and horizontal scale bar denotes the number of nucleotide substitutions per site.

doi:10.1371/journal.pmed.1002055.g001

historical Central Asia III sequences sourced from our previous global study (Fig 1B) [20]. The majority of the Central Asia III isolates carried more than three antimicrobial resistance genes, encoding resistance to a wide range of first-line antimicrobials including tetracycline (*tetA*), streptomycin (*strAB*), and co-trimoxazole (*dfrA1 and sul2*), as previously described [20]. No plasmid-mediated quinolone resistance genes (*qnr*, *qepA*, *oxqAB*, *aac(6')lb-cr*) were detected in the sequences of any of the ciprofloxacin-resistant *S. sonnei*. We additionally examined the genome sequences for mutations in the quinolone resistance-determining region (QRDR) within the DNA gyrase gene (*gyrA*) and the topoisomerase IV gene (*parC*), the regions encoding the target residues for fluoroquinolones. Overlaying these mutations on the phylogenetic tree indicated that the *gyrA*-S83L mutation, the first sequential mutation that confers reduced susceptibility against fluoroquinolones, has arisen independently within the Central Asia III clade on at least four separate occasions (Fig 1B). Amongst the isolates examined here for the first time, extensive resistance to ciprofloxacin can be attributed to a single clonal emergence event, via the sequential accumulation of *gyrA*-S83L followed by *parC*-S80I and *gyrA*-D87G, except for a single outlier isolated in Australia (Fig 1B). Strong bootstrap support at each of these nodes suggests that these mutations were sequential lineage-defining events, with the final *gyrA*-D87G mutation preceding the expansion and intercontinental dissemination of the Central Asia III clade. These three QRDR mutations were also shared by ten phenotypically uncharacterized *S. sonnei* from the US, thus providing genotypic evidence for ciprofloxacin resistance. The single outlier isolate shares the *gyrA*-S83L and *parC*-S80I QRDR mutations of the other ciprofloxacin-resistant isolates, but harbours *gyrA*-D87N rather than a *gyrA*-D87G, and is within a closely related out-group of the major ciprofloxacin-resistant clone (Fig 1B).

South Asia as a Hub of Fluoroquinolone-Resistant *S. sonnei*

We additionally mapped the country of isolation and patient travel history onto the Central Asia III phylogeny to investigate the geographical structure of the clade (Fig 1B). For the ciprofloxacin-resistant *S. sonnei* isolated from countries with a low incidence of shigellosis (Ireland, Australia, and US), and for which data on recent travel history was confirmed (27/45; 60%), India was the most commonly reported travel destination (21/27; 78%). The majority of the isolates associated with travel to India clustered closely with strains isolated in neighbouring Bhutan. These data suggest that South Asia is a primary source of ciprofloxacin-resistant *S. sonnei* that have increasingly been isolated both within and outside of Asia in recent years. Furthermore, greater genetic diversity was observed within the South Asian *S. sonnei* than within the other sampled countries (Fig 1B), suggesting that this region currently acts as the most likely geographical source population.

Our data also show evidence of regional diversification of ciprofloxacin-resistant *S. sonnei* within Asia. The phylogenetic structure is highly suggestive of a clonal expansion of ciprofloxacin-resistant *S. sonnei* in Southeast Asia, specifically within Vietnam, as indicated by a long branch with 100% bootstrap support (Fig 1B). We additionally noted that *S. sonnei* nested within this clonal expansion were also isolated from travellers returning from countries including Cambodia and Thailand, indicating that isolates from this lineage have spread widely across Southeast Asia, as well as having been introduced into Australia on at least five separate occasions. An additional well-supported subpopulation of ciprofloxacin-resistant *S. sonnei*, isolated in Ireland (five individuals with no recent history of travel and one individual returning from Germany) and the US, likely represents an extended chain of local transmission within Europe and the US (Fig 1B). Alternatively, although less likely, this subpopulation could represent multiple importations and short-term local transmission of *S. sonnei* strains originating from an unsampled source population. Whilst it was not possible to identify the geographical

source or extent of local transmission definitively, the isolates most closely related to this European/US subpopulation originated in India and Bhutan, again suggesting South Asia was the most likely origin of this subpopulation. These two expansion events in Southeast Asia and Europe/US indicate that this clone of ciprofloxacin-resistant *S. sonnei* is also capable of sustained circulation upon introduction into new locations.

Discussion

Here we provide direct evidence for the ongoing global expansion of *S. sonnei* exhibiting new and clinically relevant antimicrobial resistance profiles. What is more, as we can use phylogeography in high resolution, we can link the reservoir of these organisms to a specific region. Therefore, this study has significant implications for understanding the international trafficking of antimicrobial-resistant bacterial pathogens from Asia. Furthermore, we suggest that, as a single-serotype, human-adapted pathogen with a clonal population structure, *S. sonnei* serves as a tractable model for understanding how Gram-negative antimicrobial resistant pathogens are being regularly mobilised around the globe.

To our knowledge, this is the first study that has used whole-genome sequencing to examine the emergence and global spread of ciprofloxacin-resistant *S. sonnei*. Our data show that all sequenced extant ciprofloxacin-resistant *S. sonnei*, though sourced from disparate geographical locations, belonged to a single clonal expansion of lineage III, with South Asia being the most likely hub for its origin and spread. Our findings support previous hypotheses suggesting that ciprofloxacin-resistant *S. sonnei* in industrialised countries is being imported from South Asia [15,16]. A recent estimation of worldwide antimicrobial usage reported that India was the largest consumer of antimicrobials in 2010 [24]. Additionally, the fluoroquinolones are ranked as the most common antimicrobial prescribed for acute enteric diseases in India and Bangladesh [25,26]. The intensive use of fluoroquinolones in a region where there are foci of high population density and inconsistent access to good sanitation is likely to have contributed to emergence of ciprofloxacin-resistant enteric bacteria, such as *S. sonnei* and *Salmonella* Typhi, on the Indian subcontinent [19]. Global dissemination of these organisms is likely facilitated by the volume of travel between these regions and other areas of the world.

Our new data highlight the limitations of current typing protocols for tracking *S. sonnei*. It had been previously observed that some of the ciprofloxacin-resistant *S. sonnei* isolates in this study (originating from Bhutan and Ireland) shared a similar *Xba*I pulsed field gel electrophoresis (PFGE) pattern [14,15]. This pulsotype has been observed previously in India and Bangladesh [12,13,27–29], as well as in Canada [30], Belgium [31], and Japan [32], where the association with ciprofloxacin resistance was inconsistent. However, PFGE in this context did not offer sufficient granularity to link all of the isolates or provide sufficient resolution into the regional evolution of *S. sonnei*. Our phylogenetic analyses show that this pulsotype is associated with a phylogenetic lineage, supporting the notion that this pulsotype actually represents a widespread and pervasive subclade of Central Asia III.

This work has limitations. First, the lack of historical organisms from South Asia restricts our inference to only the current situation. Furthermore, additional contemporary organisms from other settings would have improved our understanding of the current geographical spread of this clonal group. Although the data included in this analysis was generated from organisms sampled in diverse geographic locations, the majority of sequences were retrieved from South Asia, Southeast Asia, Europe, Australia, and the Americas. It is possible, albeit unlikely, that our study might have overlooked an undersampled population that may ultimately act as an additional reservoir. This issue is inherent to all phylogeographical investigations and can potentially be overcome by global-scale prospective sampling to capture and

characterize maximal global diversity. Notwithstanding these limitations, whole-genome sequencing of these geographically disparate organisms, together with plausible epidemiological links, has provided data at the highest resolution for deciphering the emergence and international spread of ciprofloxacin-resistant *S. sonnei*. Future studies interrogating extensive spatial and temporal collections of ciprofloxacin-resistant *S. sonnei*, as well as the *S. sonnei* diversity specific to South Asia prior to and during the emergence of antimicrobial resistance, are essential to further elucidate the origins and epidemiological dynamics of these populations. These supplementary investigations will greatly aid our efforts in controlling the spread of the current ciprofloxacin-resistant clone and to prevent future emergent antimicrobial-resistant bacterial populations.

In conclusion, the international surge of ciprofloxacin-resistant *S. sonnei* clone poses a substantial global health challenge, and our data show this threat is not only manifested in sporadic cases from returning travellers but also the establishment of endemic transmission in new settings. The latter is already evident in high shigellosis incidence areas such as Southeast Asia. Therefore, integrative efforts from both the research community and public health authorities should be prioritised to track, monitor, and prevent the international spread of this key enteric pathogen.

Supporting Information

S1 Table. List of organisms used in this study for phylogenetic inference.
(XLSX)

Acknowledgments

We wish to acknowledge the staff of the Department of Enteric Disease, AFRIMS, Bangkok, Thailand for their work in isolating and maintaining the strains in this study.

Author Contributions

Conceived and designed the experiments: HCT MAR DPT NRT KEH SB. Performed the experiments: HCT DPT NDL MC MV BPH SW LB CJM TNTN DVT CNT NPHL PVV THT PS NRT KEH SB. Analyzed the data: HCT MAR DPT NDL MC MV BPH SW LB CJM TNTN DVT CNT NPHL PVV THT PT PS NRT KEH SB. Contributed reagents/materials/analysis tools: NDL MC MV BPH SW LB CJM DVT CNT NPHL PT PS NRT KEH SB. Wrote the first draft of the manuscript: HCT MAR SB. Contributed to the writing of the manuscript: HCT MAR DPT NDL MC CNT PT GT NRT KEH SB. Enrolled patients: SW LB CJM DVT CNT NPHL PT. Agree with the manuscript's results and conclusions: HCT MAR DPT NDL MC MV BPH SW LB CJM TNTN DVT CNT NPHL PVV THT PT PS GT NRT KEH SB. All authors have read, and confirm that they meet, ICMJE criteria for authorship.

References

1. Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, et al. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet. Elsevier Ltd; 2012; 379: 2151–61.
2. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet. 2013; 382: 209–22. doi: 10.1016/S0140-6736(13)60844-2 PMID: 23680352
3. Bardhan P, Faruque ASG, Naheed A, Sack DA. Decrease in shigellosis-related deaths without Shigella spp.-specific interventions, Asia. Emerg Infect Dis. 2010/10/30 ed. 2010; 16: 1718–23. doi: 10.3201/eid1611.090934 PMID: 21029529

4. Vinh H, Nhu NT, Nga T V, Duy PT, Campbell JI, Hoang N V, et al. A changing picture of shigellosis in southern Vietnam: shifting species dominance, antimicrobial susceptibility and clinical presentation. BMC Infect Dis. 2009/12/17 ed. 2009; 9: 204. doi: 10.1186/1471-2334-9-204 PMID: 20003464
5. Thompson CN, Duy PT, Baker S. The Rising Dominance of Shigella sonnei: An Intercontinental Shift in the Etiology of Bacillary Dysentery. PLoS Negl Trop Dis. 2015; 9: 1–13.
6. Bennish ML. Potentially lethal complications of shigellosis. Rev Infect Dis. 1991; 13 Suppl 4: S319–S324. PMID: 2047657
7. Vinh H, Main J, Chinh M, Tam C, Trang P, Nga D, et al. Treatment of bacillary dysentery in Vietnamese children: two doses of ofloxacin versus 5-days nalidixic acid. Trans R Soc Trop Med Hyg. 2000; 94: 323–326. PMID: 10975012
8. WHO. Guidelines for the control of shigellosis, including epidemics due to Shigella dysenteriae type 1. World Health Organization. 2005.
9. Ruiz J. Mechanisms of resistance to quinolones: Target alterations, decreased accumulation and DNA gyrase protection. J Antimicrob Chemother. 2003; 51: 1109–1117. PMID: 12697644
10. Holt KE, Vu T, Nga T, Pham D, Vinh H, Wook D, et al. Tracking the establishment of local endemic populations of an emergent enteric pathogen. PNAS. 2013; 110.
11. Horiuchi S, Inagaki Y, Yamamoto N, Okamura N, Imagawa Y, Nakaya R. Reduced susceptibilities of Shigella sonnei strains isolated from patients with dysentery to fluoroquinolones. Antimicrob Agents Chemother. 1993; 37: 2486–2489. PMID: 8285640
12. Nandy S, Dutta S, Ghosh S, Ganai A, Rajahamsan J, Theodore RBJ, et al. Foodborneassociated Shigella sonnei, India, 2009 and 2010. Emerg Infect Dis. 2011; 17: 2072–2073. 8 doi: 10.3201/eid1711.110403 PMID: 22099103
13. Ud-Din AIMS, Wahid SUH, Latif H a, Shahnaij M, Akter M, Azmi IJ, et al. Changing trends in the prevalence of Shigella species: emergence of multi-drug resistant Shigella sonnei biotype g in Bangladesh. PLoS ONE. 2013; 8: e82601. doi: 10.1371/journal.pone.0082601 PMID: 24367527
14. Ruekit S, Wangchuk S, Dorji T, Tshering KP, Pootong P, Nobthai P, et al. Molecular characterization and PCR-based replicon typing of multidrug resistant Shigella sonnei isolates from an outbreak in Thimphu, Bhutan. BMC Res Notes. BMC Research Notes; 2014; 7: 1–9.
15. De Lappe N, Connor JO, Garvey P, Mckeown P, Cormican M. Ciprofloxacin-Resistant Shigella sonnei Associated with Travel to India. Emerg Infect Dis. 2015; 21: 2013–2015.
16. Bowen A, Hurd J, Hoover C, Khachadourian Y, Traphagen E, Harvey E, et al. Importation and Domestic Transmission of Shigella sonnei Resistant to Ciprofloxacin—United States, May 2014-February 2015. Morb Mortal Wkly Rep. 2015; 64: 318–320.
17. Kim JS, Kim JJ, Kim SJ, Jeon S, Seo KY, Choi J, et al. Shigella sonnei Associated with Travel to Vietnam, Republic of Korea. Emerg Infect Dis. 2015; 21: 1247–1250. doi: 10.3201/eid2107.150363 PMID: 26079171
18. He M, Miyajima F, Roberts P, Ellison L, Pickard DJ, Martin MJ, et al. Emergence and global spread of epidemic healthcare-associated Clostridium difficile. Nat Genet.; 2013; 45: 109–13. doi: 10.1038/ng.2478 PMID: 23222960
19. Wong VK, Baker S, Pickard DJ, Parkhill J, Page AJ, Feasey N a, et al. Phylogeographical analysis of the dominant multidrug-resistant H58 clade of Salmonella Typhi identifies inter- and intracontinental transmission events. Nat Genet. 2015;
20. Holt KE, Baker S, Weill FX, Holmes EC, Kitchen A, Yu J, et al. Shigella sonnei genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. Nat Genet. 2012/08/07 ed. 2012; 44: 1056–1059. doi: 10.1038/ng.2369 PMID: 22863732
21. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25: 2078–9. doi: 10.1093/bioinformatics/btp352 PMID: 19505943
22. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008; 18: 821–9. doi: 10.1101/gr.074492.107 PMID: 18349386
23. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012; 67: 2640–4. doi: 10.1093/jac/dks261 PMID: 22782487
24. Van Boeckel TP, Gandra S, Ashok A, Caudron Q, Grenfell BT, Levin S a., et al. Global antibiotic consumption 2000 to 2010: An analysis of national pharmaceutical sales data. Lancet Infect Dis. Elsevier Ltd; 2014; 14: 742–750.
25. Fahad BM, Matin A, Shill MC, Asish KD. Antibiotic usage at a primary health care unit in Bangladesh. Australas Med J. 2010; 3: 414–421.

26. Kotwani A, Chaudhury RR, Holloway K. Antibiotic-prescribing practice of primary care prescribers for acute diarrhea in New Delhi, India. Value Heal. 2012; 15: S16–S119.

27. Pazhani GP, Niyogi SK, Singh AK, Sen B, Taneja N, Kundu M, et al. Molecular characterization of multi-drug-resistant Shigella species isolated from epidemic and endemic cases of shigellosis in India. J Med Microbiol. 2008; 57: 856–863. d doi: 10.1099/jmm.0.2008/000521-0 PMID: 18566144

28. Ghosh S, Pazhani GP, Chowdhury G, Guin S, Dutta S, Rajendran K, et al. Genetic characteristics and changing antimicrobial resistance among Shigella spp. isolated from hospitalized diarrhoeal patients in Kolkata, India. J Med Microbiol. 2011; 60: 1460–1466. 0 doi: 10.1099/jmm.0.032920-0 PMID: 21659504

29. Talukder K a., Islam Z, Dutta DK, Aminul Islam M, Khajanchi BK, Azmi IJ, et al. Antibiotic resistance and genetic diversity of Shigella sonnei isolated from patients with diarrhoea between 1999 and 2003 in Bangladesh. J Med Microbiol. 2006; 55: 1257–1263. PMID: 16914657

30. Gaudreau C, Ratnayake R, Pilon P a, Gagnon S, Roger M, Lévesque S. CiprofloxacinResistant Shigella sonnei among Men Who Have Sex with Men, Canada, 2010. Emerg Infect Dis. 2011; 17: 1747–1750. doi: 10.3201/eid1709.102034 PMID: 21888811

31. Vrints M, Mairiaux E, Van Meervenne E, Collard JM, Bertrand S. Surveillance of antibiotic susceptibility patterns among Shigella sonnei strains isolated in Belgium during the 18-year period 1990 to 2007. J Clin Microbiol. 2009; 47: 1379–1385. doi: 10.1128/JCM.02460-08 PMID: 19321731

32. Uchimura M, Kishida K, Koiwai K. Increasing Incidence and the Mechanism of Resistance of Nalidixic Acid Resistant Shigella sonnei. Kansenshogaku Zasshi. 2001; 75: 923–930. PMID: 11766375

GUT MICROBES
2017, VOL. 0, NO. 0, 1–17
https://doi.org/10.1080/19490976.2017.1361093

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER/REPORT

OPEN ACCESS Check for updates

Assessing gut microbiota perturbations during the early phase of infectious diarrhea in Vietnamese children

Hao Chung The[a], Paola Florez de Sessions[b], Song Jie[b], Duy Pham Thanh[a], Corinne N. Thompson[a,c,d], Chau Nguyen Ngoc Minh[a], Collins Wenhan Chu[b], Tuan-Anh Tran[a], Nicholas R. Thomson[d,e], Guy E. Thwaites[a,c], Maia A. Rabaa[a,c], Martin Hibberd[b,d], and Stephen Baker[a,c,f]

[a]Department of Enteric Infections, The Hospital for Tropical Diseases, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam; [b]The Genome Institute of Singapore, GIS Efficient Rapid Microbial Sequencing (GERMS), Singapore; [c]Centre for Tropical Medicine and Global Health, Oxford University, Oxford, United Kingdom; [d]The London School of Hygiene and Tropical Medicine, London, United Kingdom; [e]Infection Genomics, The Wellcome Trust Sanger Institute, Hinxton, United Kingdom; [f]The Department of Medicine, The University of Cambridge, Cambridge, United Kingdom

ABSTRACT

Diarrheal diseases remain the second most common cause of mortality in young children in developing countries. Efforts have been made to explore the impact of diarrhea on bacterial communities in the human gut, but a thorough understanding has been impeded by inadequate resolution in bacterial identification and the examination of only few etiological agents. Here, by profiling an extended region of the 16S rRNA gene in the fecal microbiome, we aimed to elucidate the nature of gut microbiome perturbations during the early phase of infectious diarrhea caused by various etiological agents in Vietnamese children. Fecal samples from 145 diarrheal cases with a confirmed infectious etiology before antimicrobial therapy and 54 control subjects were analyzed. We found that the diarrheal fecal microbiota could be robustly categorized into 4 microbial configurations that either generally resembled or were highly divergent from a healthy state. Factors such as age, nutritional status, breastfeeding, and the etiology of the infection were significantly associated with these microbial community structures. We observed a consistent elevation of *Fusobacterium mortiferum, Escherichia*, and oral microorganisms in all diarrheal fecal microbiome configurations, proposing similar mechanistic interactions, even in the absence of global dysbiosis. We additionally found that *Bifidobacterium pseudocatenulatum* was significantly depleted during dysenteric diarrhea regardless of the etiological agent, suggesting that further investigations into the use of this species as a dysentery-orientated probiotic therapy are warranted. Our findings contribute to the understanding of the complex influence of infectious diarrhea on gut microbiome and identify new opportunities for therapeutic interventions.

ARTICLE HISTORY
Received 12 June 2017
Revised 24 July 2017
Accepted 24 July 2017

KEYWORDS
diarrhea; microbiome; enterotype; developing country; Bifidobacterium; Fusobacterium; oral microbiome

Downloaded by [113.161.86.225] at 21:47 24 August 2017

Introduction

Diarrheal diseases result in approximately 1.7 billion new infections and 0.75 million deaths in children aged under 5 y annually, making it the second most common cause of mortality in young children in developing countries.[1-3] Repeated diarrheal episodes cumulatively increase the risk of malnutrition and stunting, which are associated with cognitive impairment and development of cardiovascular diseases and glucose intolerance in adulthood.[4-6] Such syndromes exhaust societal resources, especially in impoverished regions; therefore there is a substantial demand for efficacious treatments and prophylaxes.

The human gastrointestinal tract is populated with an immensely rich and diverse microbial communities, with the large intestine harboring the greatest density of bacteria.[7] Studies regarding the gut microbiota have highlighted the impact that microbial communities exert on human health, specifically in relation to nutrition,[8] metabolic diseases,[9] and cancer.[10] Various techniques, including metagenomics, have been used to investigate microbial disturbances

CONTACT Stephen Baker sbaker@oucru.org The Hospital for Tropical Diseases 764 Vo Van Kiet St., Ward 1, District 5, Ho Chi Minh City, Vietnam.
Supplemental data for this article can be accessed on the publisher's website.
H.C.T and P.F.S are joint co-first authors

© 2017 Hao Chung The, Paola Florez de Sessions, Song Jie, Duy Pham Thanh, Corinne N. Thompson, Chau Nguyen Ngoc Minh, Collins Wenhan Chu, Tuan-Anh Tran, Nicholas R. Thomson, Guy E. Thwaites, Maia A. Rabaa, Martin Hibberd, and Stephen Baker. Published with license by Taylor & Francis
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Downloaded by [113.161.86.225] at 21:47 24 August 2017

in persistent *Clostridium difficile* infections[11] and inflammatory bowel disease,[12] but such dysbiosis remains insufficiently characterized for infectious diarrhea in high incidence settings. The highly dynamic succession of microbial colonization in young children confounds analysis in this target group.[13,14] Nevertheless, several 16S rRNA gene profiling studies have shown consistent patterns in the initial gut microbiota response following acute diarrheal episodes, in which a transition toward Proteobacteria and *Streptococcus* is observed in an increasingly oxygenated environment.[15-17] This shift is additionally coupled with a reduction in specific Firmicutes and Bacteroidetes colonizers; their relative abundance is restored during post-diarrheal recovery state.[15,17-19] These previous studies have either focused on diarrhea associated with *Vibrio cholerae*,[15,18] or did not offer sufficiently detailed granularity for understanding microbiome alternation.[16,19] Here, we aimed to characterize modifications of the gut microbiota in the early phase of infectious diarrhea in Vietnamese children by examining their fecal bacterial 16S rRNA composition.

Results

16S rRNA gene sequencing of diarrheal fecal samples

We subjected a collection of 200 fecal samples (55 samples from non-diarrheal controls and 145 samples from diarrheal patients) from Vietnamese children, to DNA extraction and 16S rRNA gene sequencing (Table 1 and Table S1). 15/54 controls showed presence of at least one pathogen after detection, including Norovirus, Rotavirus, *Salmonella*, and *Campylobacter*. However, the isolation of these organisms in asymptomatic carriers is frequently observed in endemic setting.[20] One hundred and ninety nine of the fecal samples produced 16S rRNA sequences (median library size = 2,880,968 paired-end reads). Subsampling and quality filtering generated sub-libraries with a median depth of 756,829 paired-end reads, which served as the inputs for EMIRGE assemblies of the V3-V6 region. A mean of 74% of the reads (range: 37% to 92.5%) were successfully mapped and retrieved. A total of 131,702 assembled and clustered sequences, representing a pool of individual sample OTUs (operational taxonomic units), were produced from all samples. Subsequent filtering, OTU reconstruction, and chimeric sequence removal resulted in 7,479 OTUs clustering at 97% similarity in the 199 fecal samples.

Table 1. Summary of the etiologies associated with 142 diarrheal samples contributing to this study.

Etiological agent	Number	Proportion (%)
Viral	39	27.5
Norovirus (NoV)	19	13.4
Rotavirus (RoV)	20	14.1
Bacterial	80	56.3
Campylobacter	14	9.9
Salmonella	15	10.6
Shigella	49	34.5
Plesiomonas	2	1.4
Mixed	23	16.2
Campylobacter + NoV/RoV	7	4.9
Plesiomonas + NoV/RoV	1	0.7
Salmonella + NoV/RoV	4	2.8
Shigella + NoV/RoV	11	7.7
Total	142	100

Microbial structures within the fecal samples

We examined the microbial composition of the fecal specimens from controls and diarrheal cases using the relative abundance of 205 agglomerated unique genera and their phylogenetic relationships. The average relative abundance of the various OTUs in the fecal specimens, up to genus level, is shown in Fig. S1. *De novo* clustering of the weighted Unifrac dissimilarity matrix, representing the pairwise diversity between samples, segregated the samples into 4 community state types (CSTs). The optimal number of CST groups was determined by gap statistics (Fig. S2) and supported by a prediction strength of > 75%. A principal coordinate analysis (PCoA) using the weighted Unifrac matrix provided further support for the structuring of the 4 different microbial CSTs (Fig. S3). Constrained analysis on the principal coordinates of 195 samples with their associated metadata indicated that age, weight-for-age Z (WAZ) score, and disease status (diarrhea/asymptomatic) best explained the variation between CSTs (Fig. 1). A random forest algorithm was used to reclassify each sample to one of the 4 CSTs based on the relative abundance of 205 genera, yielding a predicted error rate of 11.06%. Twenty-eight genera with the highest calculated variable importance measurement (> 0.002) were selected as the most significant taxa segregating the 4 CSTs. Random forest classification was then repeated with these taxa only to re-assess the most predictive genera for CST clustering, yielding a similar misclassification error rate of

Downloaded by [113.161.86.225] at 21:47 24 August 2017

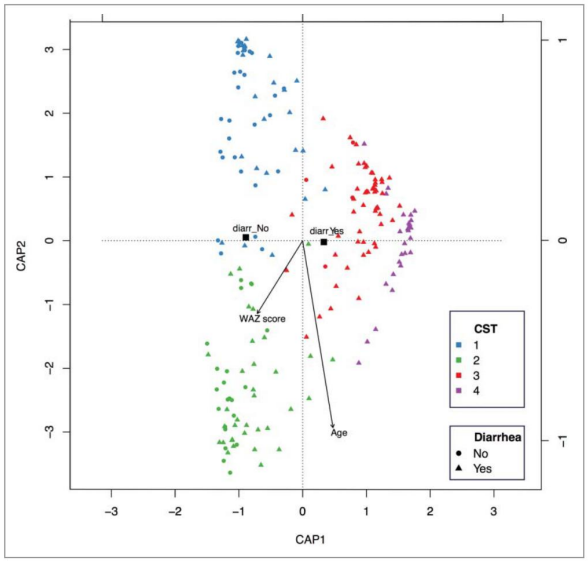


Figure 1. The impact of diarrhea and demographics on the gut bacterial configurations. Constrained analysis of principal coordinates (CAP) biplot displaying the relationship between bacterial compositions (colored circles and triangles; see legend) and selected metadata. CAP was performed on the genus-level agglomerated weighted Unifrac dissimilarity matrix of 195 fecal samples with complete metadata. Black arrows denote the magnitudes and directions of quantitative demographic variables (age and WAZ score), and black squares represent qualitative variables (centered around relative diarrhea status).

11.56% (Fig. S4A and B). The following taxa predominated in each of the CSTs (with species included based on their abundances): CST1, *Bifidobacterium*-dominant (*B. longum, B. breve, B. pseudocatenulatum, B. bifidum*); CST2, *Bacteroides*-dominant (*B. fragilis, B. vulgatus, B. uniformis, B. caccae*); CST3, *Streptococcus*-dominant (*S. gallolyticus* and *S. salivarius*); and CST4, *Escherichia*-dominant (Fig. 2). A multivariate homogeneity of group dispersions test showed that CST3 possessed the highest measurement. This high variability was explained by the fact that several samples from this CST were rich in *Megamonas*, as opposed to *Streptococcus*. For diarrheal cases, the median reported duration of diarrhea onset was 2 d (IQR: 1–3 days). Thus, we classified our findings as representing the early phase of infectious diarrhea since the gut microbiota composition differs as diarrhea progresses to later stages.[17,18]

The majority of the asymptomatic samples were within CST1 (n = 27) and CST2 (n = 22), which were generally enriched for taxa known to correspond with a healthy gut microbiota in young children.[13,21] CST2 harbored a greater degree of bacterial diversity than CST1 and encompassed various genera that delineate an increasingly mature gut microbiota, including *Faecalibacterium, Prevotella, Clostridium*,

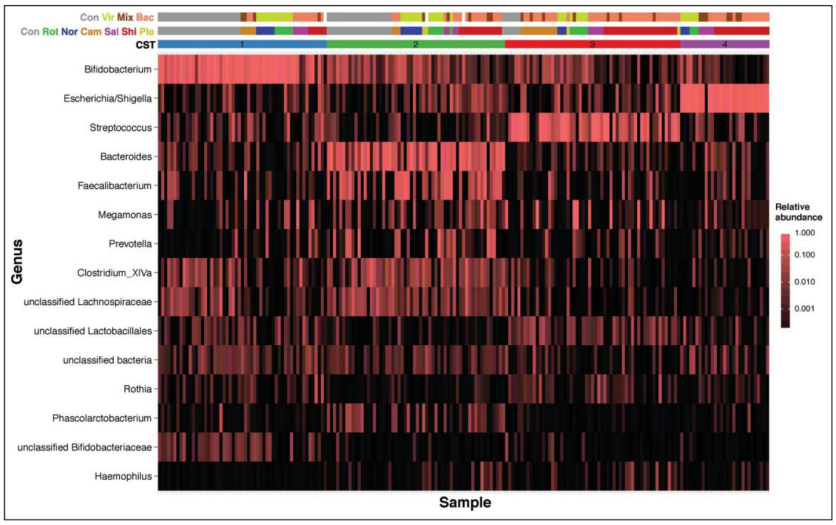


Figure 2. The bacterial compositions of the 4 community state types (CSTs). Heatmap showing the fractional abundance of the 15 most important genera determining the clustering of 199 diarrheal and non-diarrheal children fecal samples, as defined by random forest classification. Clustering was performed on the genus level weighted Unifrac dissimilarity matrix (with 205 genera) using the partitioning around meloids algorithm and identified 4 CSTs. The nature of sample is indicated at the head of the diagram: non-diarrheal control (gray); viral infection (chartreuse); mixed bacterial and viral infection (brown); bacterial infection (salmon); missing data (white). The confirmed major etiologies of infection are indicated in the middle row: non-diarrheal control (gray); Rotavirus (green); Norovirus (blue); *Campylobacter* (orange); *Salmonella* (magenta); *Shigella* (red); *Plesiomonas* (yellow).

Downloaded by [113.161.86.225] at 21:47 24 August 2017

Lachnospiraceae, and *Phascolarctobacterium*.[8,13] The fecal samples from the diarrheal cases were distributed across the 4 CSTs, with CST3 (n = 52), and CST4 (n = 29) comprised almost exclusively of diarrheal cases. This structure corresponded with the dominance of *Streptococcus* and *Escherichia* species in gut microbiota of diarrheal children, as reported previously.[16,17] The estimated Shannon diversity index was elevated in both controls and the cases in CST2 in comparison to the other CSTs ($p < 0.05$, ANOVA-Tukey's test) (Fig. 3). We additionally observed that bacterial diversity was lower in the diarrheal cases relative to the controls within CST1 and CST2, however this difference between the groups was not statistically significant ($p > 0.05$, ANOVA-Tukey's test).

Factors affecting microbial structure in the feces of children with diarrhea

The CST grouping of the samples from diarrheal children (n = 136), and their respective demographic and clinical metadata, were input into a multinomial logistic regression model to identify factors explaining the grouping of the 4 CSTs (Table 2). CST2 was selected as a reference as it represented a mature gut microbiota configuration. We found that age, WAZ score, infection type (bacterial, viral, or mixed infection (Table 1)), and feeding practices remained significantly associated with CST classification ($p < 0.05$; likelihood ratio test). Furthermore, younger age was significantly associated with CST1 (median age = 16 months; $p = 0.015$; 2-tailed Z test) and CST3 (median age = 15 months; $p = 0.042$; 2-tailed Z test) (Table S2), while the fecal samples in CST2 and CST4 were generally associated with older children (median age = 22 and 23 months, respectively). This age differential may be explained by the maturation of the gastrointestinal microbiota in children as the *Bifidobacterium*-rich stage (CST1) precedes the *Bacteroides*-Firmicutes stage (CST2).[14] Additionally, we found that a lower WAZ score was associated with CST4 (median WAZ score = −1.1; $p = 0.008$; 2-tailed Z test), while CST3 contained a significantly greater proportion of fecal samples from children infected with bacterial diarrheal pathogens (73%; $p = 0.001$; 2-tailed Z test; Table S2). Lastly, children that were exclusively breastfed were significantly associated with CST1 ($p = 0.015$; 2-tailed Z test). However, there was no significant difference in feeding patterns according to age ($p = 0.422$, Kruskal-Wallis rank sum test), suggesting that the association of breastfeeding with CST1 was not a product of clustering with younger age group.

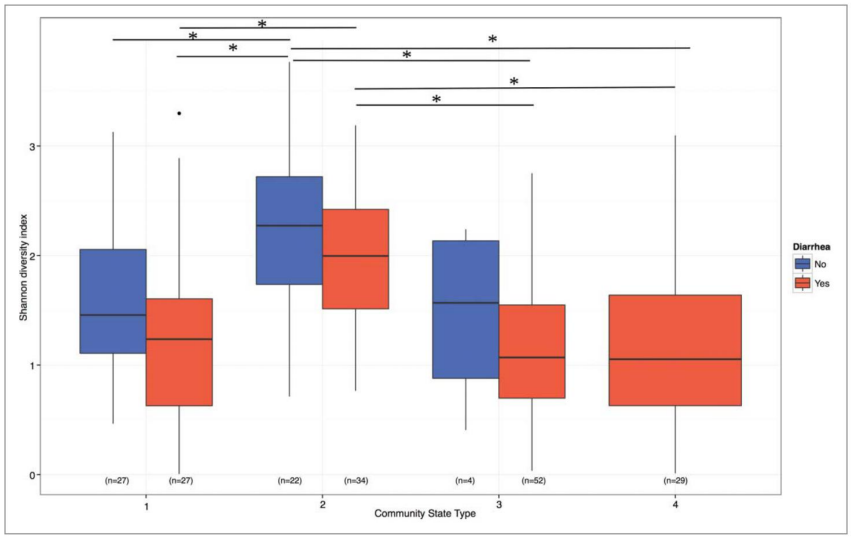


Figure 3. Shannon diversity index for 195 fecal samples, classified by diarrheal status and CST membership. Boxplots showing the Shannon diversity index, the upper whisker extends from the 75th percentile to the highest value within the 1.5 * interquartile range (IQR) of the hinge, the lower whisker extends from the 25th percentile to the lowest value within 1.5 * IQR of the hinge. Data points beyond the end of the whiskers are outliers. The asterisk indicates statistical significant in the pairwise comparison between CST2 and the other CSTs (ANOVA-Tukey's test, $p < 0.05$).

Table 2. Demographic and clinical predictors used in multinomial logistic regression for community state types (CSTs).

Patient characteristics	Odds ratio of associated factor (95% CI*); p value$		
	CST1 (*Bifidobacterium*#; N = 25) *v* CST2	CST3 (*Streptococcus*; N = 49) *v* CST2	CST 4 (*Escherichia*; N = 29) *v* CST2
Age in months	0.92 [0.86–0.98]; **$p = 0.015$**	0.95 [0.91–1.00]; **$p = 0.042$**	0.99 [0.94–1.04]; $p = 0.574$
Male	2.14 [0.62–7.41]; $p = 0.230$	2.81 [0.98–8.03]; $p = 0.054$	2.12 [0.67–6.65]; $p = 0.200$
Weight-for-age Z score	0.63 [0.35–1.12]; $p = 0.113$	0.69 [0.42–1.12]; $p = 0.135$	0.46 [0.26–0.81]; **$p = 0.008$**
Breastfed and formula-milk fed (compared with breastfed-only)	0.10 [0.01–0.64]; **$p = 0.015$**	0.81 [0.24–2.76]; $p = 0.737$	0.44 [0.11–1.71]; $p = 0.234$
Formula-milk fed only (compared with breastfed-only)	0.26 [0.06–1.21]; $p = 0.087$	1.38 [0.38–5.01]; $p = 0.621$	0.25 [0.05–1.36]; $p = 0.108$
Urban residence (compared with rural)	1.60 [0.29–8.95]; $p = 0.593$	2.17 [0.51–9.3]; $p = 0.297$	1.5 [0.33–6.8]; $p = 0.600$
Monthly income 145–483 USD (compared with monthly income < 145 USD)	1.35 [0.31–5.93]; $p = 0.690$	0.53 [0.16–1.79]; $p = 0.305$	0.48 [0.13–1.74]; $p = 0.262$
Monthly income > 484 USD (compared with monthly income < 145 USD)	3.75 [0.15–94.77]; $p = 0.422$	3.28 [0.26–41.97]; $p = 0.360$	2.42 [0.14–41.78]; $p = 0.544$
Vomiting	0.19 [0.05–0.77]; **$p = 0.020$**	0.68 [0.20–2.26]; $p = 0.525$	0.42 [0.11–1.61]; $p = 0.203$
Dysentery	0.63 [0.17–2.35]; $p = 0.496$	0.76 [0.26–2.24]; $p = 0.624$	0.56 [0.17–1.86]; $p = 0.342$
Bacterial infection (compared with viral)	2.12 [0.43–10.52]; $p = 0.358$	13.61 [3.01–61.52]; **$p = 0.001$**	4.02 [0.81–20.04]; $p = 0.089$
Mixed infection (compared with viral)	1.21 [0.19–7.78]; $p = 0.844$	5.07 [0.9–28.58]; $p = 0.066$	3.31 [0.59–18.74]; $p = 0.175$

Note. Figures in bold indicate $p < 0.05$.
*CI: Confidence Interval. $p value calculated through 2 tailed Z test.
#Bacterial taxon which is predominant in each CST

Downloaded by [113.161.86.225] at 21:47 24 August 2017

Changes in taxonomic abundance in the feces of children with diarrhea

Streptococcus and *Escherichia* species were highly abundant in the diarrheal gut communities associated with CST3 and CST4, respectively. However, 64/145 (44%) of the fecal bacterial community structures from diarrheal patients were more similar to those from the control groups (CSTs 1 and 2). We used DESeq2 to identify the OTUs that were differentially abundant in the diarrheal samples of all 4 CSTs, taking into account the overarching discrepancy in composition due to different CST membership, aiming to minimize effects associated with confounding factors such as age, WAZ scores, and infection type. Seventeen OTUs were found to be at least 4-fold higher in abundance in diarrheal samples than control samples ($p < 0.05$); the majority of these were *Escherichia* species or organisms associated with the oral cavity, including *Fusobacterium, Gemella*, and *Actinomyces* (Fig. 4A). The most significant difference in a bacterial species in diarrheal fecal samples was associated with *Fusobacterium mortiferum*, which was 2^{10} fold more abundant in diarrheal fecal samples than control samples ($p < 0.0001$). Furthermore, *Sutterella, Megamonas*, and *Enterococcus* species were also in higher abundance in diarrheal fecal samples in comparison to control samples. This analysis also identified 19 OTUs that were significantly depleted in diarrheal samples ($p < 0.05$). The majority of these OTUs (13/19) were gut colonizers belonging to the orders Clostridiales and Erysipelotrichales, including *Subdoligranulum, Roseburia, Eubacterium, Coprococcus, Catenibacterium*, and Ruminococcaceae. Among these, the most depleted taxon was the short chain fatty acid (SCFA) producer *Blautia hansenii*, which was significantly reduced in samples from those with diarrhea, regardless of the overall community group of the sample ($p < 0.0001$) (Fig. S5). *B. bifidum, B. uniformis*, and *L. reuteri* were also significantly associated with a healthy microbial configuration, albeit with different proportions in the various CSTs, likely reflecting age-specific development of the gut microbiota (Fig. S5). In addition, we performed the same analysis on a subset of samples belonging to CSTs 1 and 2, and this produced comparable results to the aforementioned.

We next constructed a correlation network that encompassed representative fecal taxa from both control and diarrheal samples. The resulting correlation network formed a tightly connecting interaction between members of the Clostridiales, Erysipelotrichales, and Bacteroidales, which indicated a positive interaction between these colonizers of a healthy gut (Fig. 5). A further positive interacting cluster was composed of genera that are frequently associated with the human oral cavity, including *Fusobacterium nucleatum, Parvimonas micra, Peptostreptococcus stomatis,*

Downloaded by [113.161.86.225] at 21:47 24 August 2017

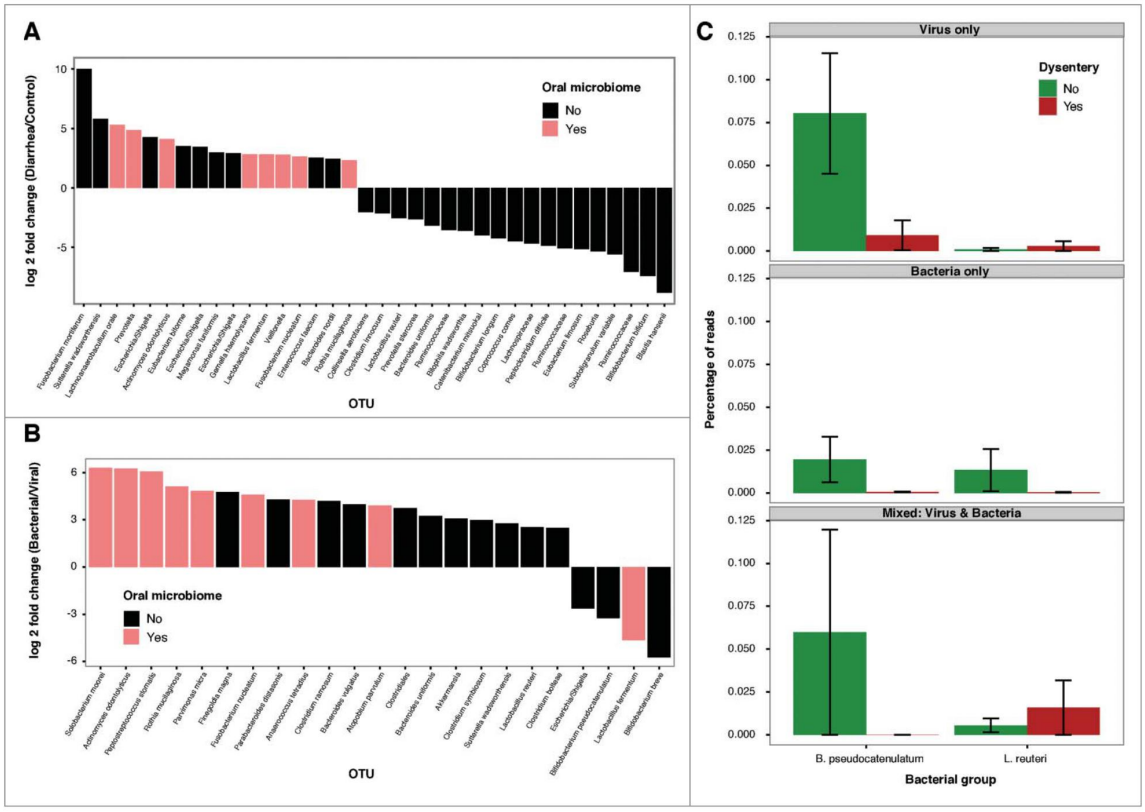


Figure 4. Bacterial taxa showing significantly different abundance among the examined classes. OTUs were identified to be of significantly differential abundance between groups in examination, as detected and filtered by DESeq2. In short, only OTUs with adjusted p values < 0.05, estimated fold change >4 or $<1/4$, and estimated base mean >30 were considered significantly differentially abundant and included in the plot. (A) Ratio of the log2 fold change of OTUs that differ between diarrheal and control stools, accounting for the different general microbial configurations in 4 CSTs (N = 199). (B) Ratio of the log2 fold change of OTUs that differ between bacterial and viral diarrheal infections, accounting for the different general microbial configurations in 4 CSTs (N = 119). (C) Relative abundances of OTUs that differ between dysenteric and non-dysenteric stools, accounting for different types of infection (N = 142).

Solobacterium moorei, Gemmella haemolysans, Actinomyces odontolyticus, Lachnoanaerobaculum orale, Abiotrophia defectiva, Atopobium parvulum, Rothia mucilaginosa, and *Streptococcus salivarius* (Fig. 5). Conversely, *Rothia* was negatively correlated with *Bacteroides vulgatus* and *Faecalibacterium prausnitzii*, which are both indicators of a healthy and mature gastrointestinal microbiota.[8,13] We additionally found that *Fusobacterium mortiferum*, the significantly enriched taxon in fecal samples from diarrheal children, exhibited a negative association with a key component of the gut Clostridiales network (*Clostridium innocuum*). This suggests that the colonization of *F. mortiferum* may inhibit the proliferation of related dependent taxa such as *Blautia, Ruminococcus*, and several *Bacteroides*. The negative correlation observed between *Bifidobacterium longum* subsp. *infantis* and *Bifidobacterium breve* likely reflects a competitive ecology as the gut microbiota matures.[22]

Taxonomic changes associated with bacterial induced diarrhea and dysentery

To explore the differential composition of the gut microbiota in response to the different etiological agents (viral vs. bacterial), we examined a subset of 119 diarrheal fecal samples (viral infections = 39; bacterial infections = 80) to eliminate potential misclassification induced by 23 cases of mixed infections. Analysis using DESeq2 identified a group of 19 OTUs that were significantly more abundant in bacterial infections across the 4 CSTs. Among these, 8 taxa were more commonly associated with human oral microbiota (Fig. 4B); *Actinomyces* and *Rothia* were the most prevalent (assessed

Downloaded by [113.161.86.225] at 21:47 24 August 2017

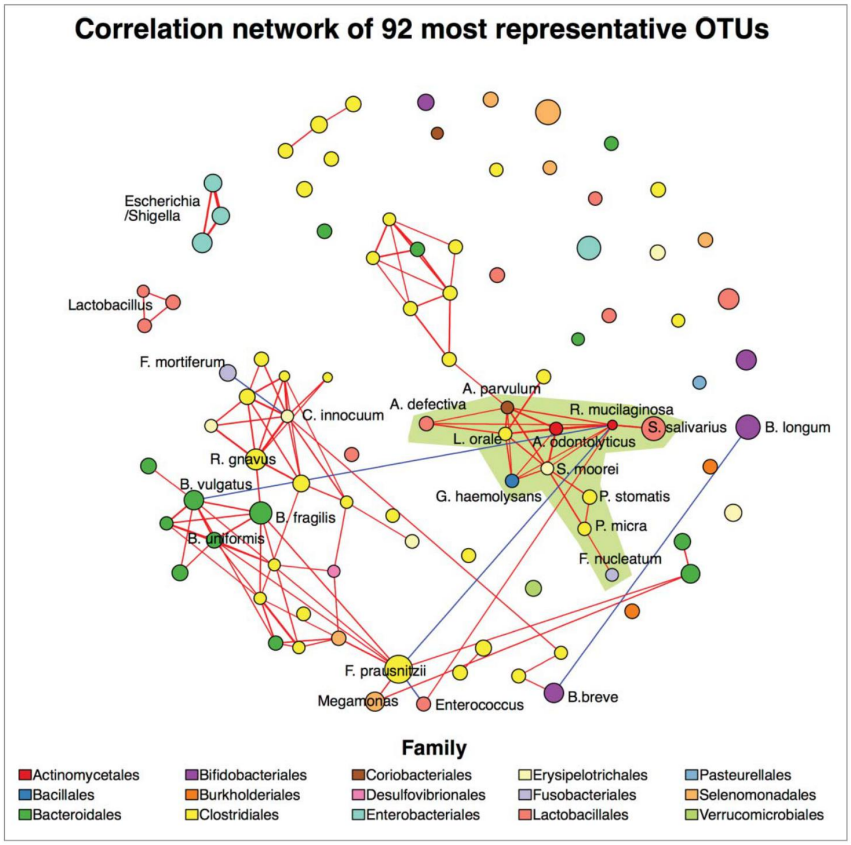


Figure 5. Correlation network of the healthy and diarrheal microbiome. Figure shows the correlation network of the 92 most represented OTUs sampled from 199 stool isolates, defined as OTUs with occurrence in at least 10 samples, constructed using the SparCC wrapper from package 'SpiecEasi'. Only correlations with calculated p value ≤ 0.05 and absolute magnitude ≥ 0.25 were shown in the network. Positive and negative interactions were denoted as a red and blue solid line respectively, with line weight proportional to correlation strength. The OTUs (nodes) were colored based on taxonomic family (see legend), with sizes proportional to their relative abundances. The light green shaded area covers OTUs identified as members of normal human oral microbiota (through comparison with the Human Oral Microbiome Database).

by the DESeq2 calculated base mean value). Other organisms with high relative proportion in the fecal samples from children with bacterial induced diarrhea included *C. ramosum* and *B. vulgatus*. Conversely, 4 OTUs were significantly elevated in the viral diarrheal infections; these included the *Bifidobacterium* species, *B. breve* and *B. pseudocatenulatum* (Fig. 4B).

We lastly categorized 142 diarrheal fecal samples based on the diagnosis of dysentery (the visible presence of mucoid and/or blood in the stool; n = 60), accounting for the differential microbiota changes due to types of infectious agents (bacterial, viral, or mixed infection). We identified 2 OTUs that were significantly associated with fecal samples from non-dysenteric diarrhea (DESeq2, $p < 0.05$) (Fig. 4C). *B. pseudocatenulatum* was almost 2^5 fold less abundant in dysenteric feces in comparison to non-dysenteric diarrheal feces; this was consistent across all of the corresponding etiological agents (Fig. 4C). For bacterial infections, *L. reuteri* was uniformly more depleted in the examined dysenteric fecal samples, but its variability was higher in dysenteric stools in those with a viral or mixed infection. When we accounted only for the stool with visible blood (n = 10), the analysis revealed that *S. gallolyticus* was the most depleted taxon in dysentery compared with watery diarrhea (DESeq2 Wald test without effect of infection types, $p < 0.05$). Other taxa of decreased abundance included *Escherichia, B. pseudocatenulatum, Enterococcus* and several oral microbiome species. However, these interpretations are to be confirmed in larger studies.

Downloaded by [113.161.86.225] at 21:47 24 August 2017

Discussion

Here we assessed the impact of diarrheal infection on the composition of the gut microbiota in Vietnamese children. Our results predict that the gut microbiota in young children exhibits differential responses during the early phase of diarrhea, which can be grouped into microbial community structures that either closely resemble or are highly divergent from those of healthy children. Approximately 44% of the diarrheal cases in this study retained a *Bifidobacterium*- or *Bacteroides*-rich structure. This association largely resembles the well-studied age-dependent normal gut microbiota in children,[14] outlining a lack of global dysbiosis in these children. Among the various perturbed microbial states, we found that age, nutritional status, breastfeeding, and diarrheal etiology contributed to the composition of the bacterial communities during the early phase of diarrhea. However, since the WAZ score was collected after the onset of diarrhea, weight could be subjected to temporary fluctuation and its surrogate for nutritional status should be interpreted with caution. A recent study on diarrheal microbiome from U.S patients reported that diarrheal microbial structures could be grouped into 4 major clusters irrespective to the etiologies, of which 2 are almost exclusively restricted to diarrheal patients and associated with higher prevalence of *Escherichia*.[23] In our study, *Streptococcus* was associated with younger age and bacterial infection, while *Escherichia* was associated with older children and a poor nutritional status. Indeed, analysis of the microbiota in diarrheal samples from young children (<2 y old) predominantly infected with pathogenic *E. coli* revealed that *Streptococcus* (*S. gallolyticus* and *S. salivarius*) was the most abundant organisms during the early phase of diarrhea.[24] These 2 species were also dominant in our study. In contrast, *Escherichia* were overrepresented in the fecal samples from children (2–3 y old) infected with cholera in Bangladesh.[15] Due to limitations of the study, we were unable to assess how these initial differential microbial communities affect diarrheal severity and recovery. Dysentery and antimicrobial treatment may further confound these associations. However, the ratio of *Streptococcus*/*Bifidobacterium* showed a significant positive correlation with diarrheal fecal output and duration of hospitalization, suggesting that the higher degree of divergence from the healthy microbiota may predict a more severe clinical presentation.[24] Here, for bacterial induced diarrhea cases, we attempted to identify the pathogens' attributed sequenced reads in the fecal samples. However, only one *Salmonella* and one *Campylobacter* infections showed detectable 16S rRNA sequences classified to the corresponding pathogens. *Shigella* infections were not considered since its sequences are indistinguishable from *Escherichia*. This concurs with previous finding that by using 16S rRNA profiling on cholera fecal samples, *V. cholerae* is only detectable during the first day of symptom onset, and its presence is significantly reduced even on the second day of diarrhea.[17]

We additionally observed that obligate anaerobes belonging to the orders Clostridiales and Erysipelotrichales were consistently depleted in the diarrheal cases in comparison to the controls. This group included multiple SCFA producers including *Blautia, Subdoligranulum, Roseburia*, and *Eubacterium*. This finding may explain the transient depletion of SCFA following diarrhea, resulting in poor water and electrolyte absorption and a shortage of metabolic energy for enterocytes, which may lead to dehydration and further complications.[25] These data are in accordance with previous studies, which have shown that Clostridiales become gradually enriched during the diarrheal recovery period.[15,17,18] This loss of obligate anaerobes is coupled with the proliferation of facultative anaerobes, such as the *Streptococcus* and *Escherichia* species.[16,17] Unpredictably, we observed that a substantial component of the microbiota from the human oral cavity was consistently represented in the diarrheal fecal samples. We propose that the transiently oxygenated environment and potential decrease in bacterial competition in the diarrheal gut enhances the colonization potential of these species.[17] This may explain why *S. salivarius*, a dominant taxon of the oral cavity of young children, is capable of establishing an exceptionally high rate of colonization in suckling infants. Oral commensal organisms, predominantly *S. salivarius* and *F. nucleatum*, may serve as nucleators for the adherence of other bacteria, eventually forming polymicrobial biofilms that are capable of surviving mechanical and chemical stresses during their passage through the large intestine.[26,27] Alternatively, the diarrhea-associated obligate anaerobe, *F. mortiferum*, which was found not to be associated with oral organisms, exhibited antagonism with the Clostridiales. This finding is in agreement with the spike of

Downloaded by [113.161.86.225] at 21:47 24 August 2017

Fusobacterium observed in early phase diarrhea.[18,19] Unlike the majority of *Fusobacterium, F. mortiferum* is extremely pleomorphic and bile resistant, with the ability to rapidly accumulate and metabolize a broad spectrum of sugars independent of amino acid fermentation.[28-30] Furthermore, poultry isolated *F. mortiferum* has been shown to secrete a bacteriocin that can inhibit the growth of Gram-positive *Bacillus* species.[31] Taken together, this evidence implies that *F. mortiferum* is more resilient to environmental stresses and able to occupy the anaerobic niche of the gut, possibly accelerating the depletion of Clostridiales via bacteriocin-mediated competition.

Dysentery denotes a more severe form of diarrhea, frequently requiring antimicrobial treatment and resulting in longer periods of hospitalization.[20,32] To date, only *Lactobacillus ruminis* has been found to be depleted in dysenteric fecal samples.[16] Here, we found that *L. reuteri* was significantly diminished in dysenteric diarrhea; we speculate that this discrepancy in species level identification may be induced by variation in environmental and genetic factors of the studied populations. In addition, we report here that a *Bifidobacterium* species, *B. pseudocatenulatum*, is substantially depleted in fecal specimens following bacterial infection; this effect is amplified during dysentery. Previous research examining the V1-V2 region of the 16S rRNA failed to efficiently identify *Bifidobacterium*, thus discounting its contribution.[16] The oral administration of live *B. pseudocatenulatum* in induced cirrhotic and diabetic mice has been shown to restore the gut integrity and initiate an anti-inflammatory cytokine profile in the intestine, leading to a diminished systematic inflammatory response.[33-35] Further, this potential probiotic also induces monocyte-derived macrophages to undergo transition into the anti-inflammatory M2 phase.[36] It has been shown that the colonization of *B. pseudocatenulatum* or *B. breve*, but not other *Bifidobacterium*, protect mice from lethal Shiga toxin producing *E. coli* (STEC) infections.[37] Treatment for dysentery is currently focused on antimicrobial therapy, and the usage of broad spectrum antimicrobials quickly depletes the balanced microbiota derived metabolites such as SCFAs, leading to disruption of the gut immune homeostasis.[38] Therefore, we propose that *B. pseudocatenulatum* may be a potential dysentery-oriented probiotic candidate to reduce inflammation associated pathological conditions and accelerate the recovery of the microbiota to a healthy state.

Some limitations should be considered in the context of our investigation. First, the EMIRGE mediated read recovery rate was consistently lower in diarrheal samples. This means that a proportion of uncharacterized taxa were unavailable for examination, and it may impact on the true estimation of the diversity within these samples. Second, an exhaustive search for other diarrheagenic etiologies by molecular methods was not possible during this study; this may underestimate the true incidence of mixed infections. We propose that more comprehensive approaches, such as the Luminex xTAG multi-pathogen panel assay, could be incorporated in further studies to improve diagnostic accuracy.[39] Third, the non-diarrhea controls included are not ideally representative of the healthy population, given that some of these children sought medical help for nutritional concerns. Notwithstanding these limitations, through sequencing an extended region of the 16S rRNA gene, it was possible to obtain high confidence taxonomic assignment up to the species level, portraying the highly complex gut microbiota in greater resolution. The inclusion of various diarrheal etiologies and the statistical methods used here further permitted us to reveal the general impact of infectious diarrhea on the gut microbiota. Diarrhea cases with unknown etiologies were excluded from this study, and this may underrepresent the true heterogeneity of diarrheal causes and their differing influences on the microbiota. Furthermore, we cannot dismiss that additional underlying mechanisms such as inflammation induced changes in immune response and oxygen gradient,[40] aside from the physical manifestation of diarrhea, may contribute to observed changes in the microbiota. Regarding the correlation network analysis, it is apparent that SparCC detects fewer negative (competitive relationships) than positive correlations (mutualism and commensalism). Unlike positive correlations, which are more readily associated with the co-occurrence of OTUs, negative correlations are typically inferred by the mutual exclusions of OTUs. SparCC has been shown to possess increased precision in detecting negative correlations in comparison to other methods, but correlation detection is limited to those with greater power and low noise.[41]

In conclusion, by studying the microbiota of children with diarrhea associated with a diverse range

of confirmed etiological agents, we have generated a comprehensive insight into the effects of infectious diarrhea on the gut microbiota. Therefore, our study provides a pivotal understanding of the impact of infectious diarrhea on the gut microbiota of children, particularly ones in low-middle income urban settings in Southeast Asia. Future work, through longitudinal study designs and the employment of shotgun metagenomics, should address how the initial differential responses in gut bacterial composition may impact disease progression, such as diarrhea clearance and recovery rate of the depleted microbiota.

Materials and methods

Study design, sample collection and microbiological procedures

Samples in this study originated from a previously described prospective observational study of pediatric diarrhea conducted at 3 major hospitals in Ho Chi Minh City (HCMC), Vietnam: Children's Hospital 1 (CH1), Children's Hospital 2 (CH2), and the Hospital for Tropical Diseases (HTD).[20] The Oxford Tropical Research Ethics Committee (OxTREC) and the Scientific and Ethics Committed of CH1, CH2, and HTD provided ethical approval for the primary and this subsequent analysis. The general inclusion criteria for both diarrheal cases and non-diarrheal controls were children aged under 5 years, residing within HCMC and reporting no antimicrobial usage within 3 d before hospital admission. Children admitted to the study sites with diarrhea (defined as 3 or more loose stools or at least one bloody loose stool within a 24 hour period) were included as cases. Children who presented for health checks, nutritional, or gastrointestinal issues but reported no diarrheal or respiratory illnesses within 7 d of admission were recruited as controls. For each enrollee, clinical data regarding the symptoms and duration of diarrhea was obtained from a case report form completed by study clinicians, while demographic, feeding behavior and socioeconomic details were provided through a confidential questionnaire. Weight-for-age Z (WAZ) score was used to evaluate the nutritional status of all enrolled children based on WHO standards.[42]

A fecal sample was collected from each participant, both cases and controls, before any prescribed antimicrobial treatment. Both case and control fecal samples were subjected to standard microbiological culturing and biochemical testing to identify common diarrheal bacteria, including *Shigella, Salmonella, Campylobacter, Plesiomonas, Yersinia*, and *Aeromonas*, as described previously[43] (https://wwwnc.cdc.gov/EID/article/19/6/11–1862-Techapp1.pdf). For the detection of Rotavirus and Norovirus, reverse transcriptase polymerase chain reaction (RT-PCR) was performed on total viral RNA extracted from all specimens.[44] Further detection of *Giardia lamblia, Entamoeba histolytica*, and *Cryptosporidium* cysts were performed by microscopy on fresh fecal samples diluted in phosphate buffered saline. Stool samples were stored at −80°C until being subjected to DNA extraction up to 6 months after initial storage.

DNA extraction and sequencing

The prospective study from which the fecal specimens originated successfully recruited 1,419 cases and 609 controls. A subset of 200 fecal samples was selected for this investigation; these included 55 randomly selected controls and 145 cases from multiple confirmed etiologies. For diarrheal cases, random subsamples from each identified etiology were included to represent the epidemiology reported in our previous prospective study,[20] with emphasis on *Shigella* as it was the major cause of dysentery (Table 1). Total DNA was extracted from these fecal samples using a phenol-chloroform extraction method. Briefly, 200 μl of the fecal sample was suspended in a solution containing 500 μl of DNA extraction buffer, 50 μl NaCl 200mM, 50 μl SDS 10%, 500 μl of phenol:chloroform:isoamyl (25:24:1), and 0.7 g of 0.5-mm-diameter zirconia/silica beads. Cells were lysed by mechanical disruption with a bead beater for one minute and subjected to 3 rounds of phenol: chloroform extraction. DNA was resuspended in TE buffer supplemented with RNase.

All DNA samples were shipped on dry ice to the Genome Institute of Singapore (GIS) for 16S rRNA sequencing using the GIS Efficient Rapid Microbial Sequencing (GERMS) platform. To obtain high resolution taxonomic identification up to the species level, all samples were PCR amplified using a previously optimized primer set (338F-1061R), which produces >700 bp amplicons covering the 4 variable regions (V3-V6) of the 16S rRNA region and can be assembled *in silico* to retrieve > 92% of the sequences in the Greengenes database.[45] Long PCR fragments were cleaned using 1X AMPURE beads and randomly

Downloaded by [113.161.86.225] at 21:47 24 August 2017

Downloaded by [113.161.86.225] at 21:47 24 August 2017

fragmented with Covaris (model LE220) shearing to ~200bp. Library preparation was performed using the GeneRead DNA Library I Core Kit (Qiagen, Germany) according to the manufacturer's instructions, and the prepared DNA was subjected to sequencing on an Illumina HiSeq2500 platform to produce 75 bp paired-end reads.

EMIRGE assembly of 16S rRNA amplicons and OTU clustering

EMIRGE (Expectation Maximization Iterative Reconstruction of Genes from the Environment) is a database dependent assembler built on an iterative expectation-maximization (EM) algorithm, which is used to reconstruct full length 16S rRNA small subunit gene sequences (by simultaneously mapping and clustering) and estimate their relative abundance.[46] The SILVA small subunit (SSU) rRNA database version 119 was filtered to remove potential large subunit (LSU) rRNA sequences, and closely related sequences were clustered at 97% identity by USEARCH; this was achieved by using PhyloFlash v2.0 (https://github.com/HRGV/phyloFlash). The resulting database was used as the reference template for EMIRGE. To limit the computational effort related to using the complete data set, one million paired-end reads were randomly subsampled without replacement from each sample library using seqtk (https://github.com/lh3/seqtk). Reads were trimmed using Sickle to remove those with quality <30 and length <60.[47] We inputted the trimmed reads from each subsampled library into the amplicon-optimized version of EMIRGE, with 40 iterations and a 97% joining threshold. The EMIRGE output, a set of assembled and clustered sequences, for each sample indicates its representative OTUs with their estimated abundances. Assembled sequences with sample-wise normalized relative abundances less than 0.01% were removed from further analysis.

A pseudo-count for each sequence per subsampled library was calculated by scaling the number of successfully mapped reads to the EMIRGE estimated relative abundance of each sequence. All filtered sequences and a count table detailing their respective abundance were pooled and imported into the 16S rRNA processing platform mothur v.1.36.0.[48] To minimize the length differences in assembled sequences and to facilitate more accurate OTU clustering, sequences were aligned to a trimmed version of the SILVA reference database to include only the amplified region (338F-1061R). Gap-only columns were filtered from this alignment, and sequences with a maximum of 7 (~1%) ambiguous sites were retained for downstream analyses. Sequences were dereplicated, and ambiguous sites were replaced randomly with one of the 4 nucleotides (ATCG). These unique sequences and their respective counts served as input for the UPARSE clustering algorithm, which clustered all pooled sequences at a 97% similarity threshold.[49] Chimeric sequences were stringently predicted and removed by UCHIME v6.0, set at both *de novo* mode and against the ChimeraSlayer informed reference database (e.g. 'gold' database). A total of 7,479 OTUs were reconstructed for 199 of the successfully sequenced samples. An alignment consisting of the most abundant representative sequences from each OTU was used to construct a phylogenetic tree using FastTree 2 under default parameters.[50] Taxonomic assignments of OTU representatives up to the genus level were performed using the mothur implemented Ribosomal Database Project (RDP) classifier, with a minimum support threshold of 80%.[51]

Data analysis

All analyses were conducted in R[52] using multiple packages, including 'phyloseq', 'cluster', 'randomForestSRC', 'ggplot2', 'nnet', 'lmtest', 'vegan', 'DESeq2', 'SpiecEasi' and other packages.[53-61] An OTU count table, taxonomy classification table, related clinical and demographic data and the OTU phylogenetic tree were imported and analyzed as a 'phyloseq' object, allowing a unified and interactive analysis approach.[53]

Clustering into community state types (CSTs)

To gain overall insight into the microbial compositional structure of both controls and diarrheal cases, we applied a previously described clustering approach.[62] To reduce sparsity as well as account for the divergence and the functional similarity shared between members of the same genus, which is frequently reported for human gut microbiota, the OTU count table was agglomerated at the unique genus level for clustering as recommended.[63] Recent benchmark studies have proposed that the weighted Unifrac distance produces desirable accuracy and power in exploring β-diversity.[64-66] A weighted Unifrac dissimilarity matrix between all samples was calculated using

Downloaded by [113.161.86.225] at 21:47 24 August 2017

the collapsed genera's relative abundances and their phylogenetic relatedness.[67] The partitioning around meloids (pam) algorithm was applied to this matrix to cluster all samples into distinct CSTs, with the optimal number of CSTs (k = 4) determined using gap statistics, a goodness of clustering measure.[68] Further assessments such as prediction strength, average silhouette width (asw) were used to validate this optimal number of clusters, as recommended previously.[63] Microbial CST grouping has been applied extensively to profile the human vaginal microbiome and has produced consistent findings.[62,69,70] To assess the performance of this clustering approach and identify the most important genera contributing to the separation of these CSTs, the random forest classification algorithm was applied to the genera relative abundance table, using the samples' CST memberships as the response variable.[71] A single tree assembly (5,000 trees) was used for all 199 samples of 4 CSTs, and other parameters were set as default according to the 'rfsrc' function in package 'randomforestSRC'.[55]

Statistical analysis and regression modeling

To examine the relationship between the microbiome structures and associated explanatory variables for 195 samples with complete metadata, we applied constrained analysis of principal coordinates (CAP) on the calculated weighted Unifrac dissimilarity matrix and a set of demographic variables (age, sex, WAZ score, feeding pattern, income, rural residence), as well as diarrheal status. This was performed using the 'capscale' function in the 'vegan' package.[59,72] Significant variables were identified and included in the final model based on Akaike information criterion (AIC) in a stepwise model selection approach. The α diversity of these 195 samples was estimated by Shannon diversity index. Analysis of variance (ANOVA) with post-hoc Tukey test and Bonferroni correction for multiple comparisons was used to compare control and diarrheal α-diversity among and within each CST. Multinomial logistic regression modeling was applied using the 'multinom' function in 'nnet' package to evaluate the association of the various aforementioned demographic factors and clinical features (vomiting, dysentery, and infection type) to the CST membership of diarrheal cases, with the Bacteroides rich CST (CST2) serving as the reference group.[57] These predictors were included on the basis of limited missing data and capable of being assessed by clinicians upon patient's admission. Five samples with WAZ scores >3 or <-3, and one CST1 sample with age exceeding 2 standard deviations were considered as outliers and removed, resulting in 136 diarrheal samples with full metadata being subjected to regression modeling.

Evaluating differential abundance

DESeq2 was used to normalize the 7,479-OTUs table and to detect OTUs that show significantly differential relative abundance between 2 assessed groups per comparison (diarrhea against controls, bacterial infection against viral infection, and dysentery against non-dysentery), as recommended in recent benchmark studies.[60,64,65] To identify consistent trends observed across all CSTs (diarrhea against controls, bacterial infection against viral infection) or infection types (dysentery against non-dysentery), we applied the likelihood ratio test (LRT) approach with the reduced models incorporating only either CST or infection type as coefficients. OTUs with adjusted p values < 0.05, estimated fold change >4 or <1/4, and estimated base mean >30 were considered significantly differentially abundant between 2 examined classes. Taxonomic identification to the species level was performed by manually comparing the reconstructed 16S rRNA sequences of OTUs to either the Human Oral Microbial Database (HOMD; http://www.homd.org/) or the Human Microbiome Project (HMP; http://www.hmpdacc.org/resources/blast.php) by BLAST. OTU was assigned a species nomenclature if its sequence is of at least 97% similarity to that in the database. Using these approaches, *Shigella* sequences are indistinguishable from *Escherichia*.

Interaction network construction

A correlation network was constructed for all 199 control and diarrheal samples to characterize the potential interactions between most representative 92 OTUs, defined as OTUs detected in at least 10 samples. This filtering step did not substantially affect the representativeness of the data set, with the median sample retainment rate of 93% (IQR: 87% - 97%). The network was constructed using the SparCC wrapper in the package 'SpiecEasi'.[61,73] The statistical significance for each interaction was assessed by 100 bootstrap iterations, with p values adjusted for multiple comparison correction. To avoid spurious correlations, only

those with adjusted p values no greater than 0.05 and absolute magnitude equal to or above 0.25 were considered as significant correlations and represented in the final plots.

Abbreviations

AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
CST	Community State Type
CAP	Constrained analysis of principal coordinates
EMIRGE	Expectation Maximization Iterative Reconstruction of Genes from the Environment
HCMC	Ho Chi Minh City
IQR	Interquartile range
LRT	Likelihood Ratio Test
OTU	Operational Taxa Unit
SCFA	Short Chain Fatty Acid
WAZ	Weight for Age Z score

Ethical approval

The Oxford Tropical Research Ethics Committee (OxTREC) and the Scientific and Ethics Committee of Ho Chi Minh city's Children Hospital 1 (CH1), Children Hospital 2 (CH2), and Hospital for Tropical Diseases (HTD) provided ethical approval for the primary and this subsequent analysis (OxTREC No. 0109).

Reagents and catalog number

GeneRead DNA Library I Core Kit (Qiagen, catalog no. 180434)

Disclosure of potential conflicts of interest

The authors report no potential conflict of interest.

Acknowledgements

We gratefully acknowledge Andreas Sundquist and Matthew Davis, who wrote and modified the C++ script that generated the plot in Supplementary Figure 1. Computational resources were funded by the Li Ka Shing – University of Oxford Global Health Program (LG05, SM27). We express our gratitude to Paul J. McMurdie, Susan P. Holmes and her team for developing the package 'phyloseq' and conducting invaluable benchmark studies on analysis of microbiome data.

Funding

This work was funded by the Wellcome Trust and the Genome Institute of Singapore. SB is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (100087/Z/12/Z). DPT is supported by an OAK foundation fellowship.

ORCID

Hao Chung The http://orcid.org/0000-0002-4028-4074

References

[1] Fischer Walker CL, Perin J, Aryee MJ, Boschi-Pinto C, Black RE. Diarrhea incidence in low- and middle-income countries in 1990 and 2010: a systematic review. BMC Public Health [Internet]. 2012 [cited 2014 Nov 18];12:220. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3323412&tool=pmcentrez&rendertype=abstract. doi:10.1186/1471-2458-12-220

[2] Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, Rudan I, Campbell H, Cibulskis R, Li M, et al. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet [Internet]. 2012 [cited 2014 Jul 10];379:2151-61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22579125. doi:10.1016/S0140-6736(12)60560-1

[3] WHO. World Health Statistics 2014. Geneva, Switzerland; 2014.

[4] Checkley W, Buckley G, Gilman RH, Assis AM, Guerrant RL, Morris SS, Molbak K, Valentiner-Branth P, Lanata CF, Black RE. Multi-country analysis of the effects of diarrhoea on childhood stunting. Int J Epidemiol. 2008;37:816-30. doi:10.1093/ije/dyn099 PMID:18567626

[5] Guerrant RL, DeBoer MD, Moore SR, Scharf RJ, Lima AAM. The impoverished gut–a triple burden of diarrhoea, stunting and chronic disease. Nat Rev Gastroenterol Hepatol [Internet]. 2013;10:220-9. Available from: https://doi.org/10.1038/nrgastro.2012.239. doi:10.1038/nrgastro.2012.239

[6] Deboer MD, Lima AAM, Oría RB, Scharf RJ, Moore SR, Luna MA, Guerrant RL. Early childhood growth failure and the developmental origins of adult disease: Do enteric infections and malnutrition increase risk for the metabolic syndrome? Nutr Rev. 2012;70:642-53. doi:10.1111/j.1753-4887.2012.00543.x PMID:23110643

[7] The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature [Internet]. 2012 [cited 2013 Nov 6];486:207-14. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3564958&tool=pmcentrez&rendertype=abstract. doi:10.1038/nature11234

[8] Subramanian S, Huq S, Yatsunenko T, Haque R, Mahfuz M, Alam MA, Benezra A, DeStefano J, Meier MF, Muegge BD, et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. Nature [Internet]. 2014;510:417-21. Available from: http://www.nature.com/doifinder/10.1038/nature13421.

[9] Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome

Downloaded by [113.161.86.225] at 21:47 24 August 2017

Downloaded by [113.161.86.225] at 21:47 24 August 2017

with increased capacity for energy harvest. Nature [Internet]. 2006;444:1027-31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17183312. doi:10.1038/nature05414

[10] Kostic AD, Gevers D, Pedamallu CS, Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, Ojesina AI, et al. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Res. 2012;22:292-8. doi:10.1101/gr.126573.111 PMID:22009990

[11] Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, No D, Liu H, Kinnebrew M, Viale A, et al. Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. Nature [Internet]. 2014;517:205-8. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4354891&tool=pmcentrez&rendertype=abstract. doi:10.1038/nature13828

[12] Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward D V, Reyes JA, Shah SA, LeLeiko N, Snapper SB, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol [Internet]. 2012;13:R79. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3506950&tool=pmcentrez&rendertype=abstract. doi:10.1186/gb-2012-13-9-r79

[13] Bäckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, Li Y, Xia Y, Xie H, Zhong H, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe [Internet]. 2015;17:690-703. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25974306. doi:10.1016/j.chom.2015.04.004

[14] Dogra S, Sakwinska O, Soh S, Ngom-bru C, Brück M, Berger B, Brüssow H, Karnani N, Lee YS, Yap F, et al. Rate of establishing the gut microbiota in infancy has Consequences for Future Health. Gut Microbes. 2015;6:321-5. doi:10.1080/19490976.2015.1078051 PMID:26516657

[15] Monira S, Nakamura S, Gotoh K, Izutsu K, Watanabe H, Alam NH, Nakaya T, Horii T, Ali SI, Iida T, et al. Metagenomic profile of gut microbiota in children during cholera and recovery. Gut Pathog [Internet]. 2013;5:1. Available from: http://www.gutpathogens.com/content/5/1/1. doi:10.1186/1757-4749-5-1

[16] Pop M, Walker AW, Paulson J, Lindsay B, Antonio M, Hossain M, Oundo J, Tamboura B, Mai V, Astrovskaya I, et al. Diarrhea in young children from low-income countries leads to large-scale alterations in intestinal microbiota composition. Genome Biol [Internet]. 2014;15:R76. Available from: http://genomebiology.com/2014/15/6/R76. doi:10.1186/gb-2014-15-6-r76

[17] David L, Weil A, Ryan ET, Calderwood SB, Harris JB, Chowdhury F, Begum Y, Qadri F, Larocque RC, Turnbaugh J. Gut Microbial Succession Follows Acute Secretory Diarrhea in Humans. MBio. 2015;6:1-14. doi:10.1128/mBio.00381-15

[18] Hsiao A, Ahmed AMS, Subramanian S, Griffin NW, Drewry LL, Petri WA, Haque R, Ahmed T, Gordon JI. Members of the human gut microbiota involved in recovery from Vibrio cholerae infection. Nature [Internet]. 2014;515:423-6. Available from: http://www.nature.com/doifinder/10.1038/nature13738. doi:10.1038/nature13738

[19] Becker-dreps S, Allali I, Monteagudo A, Vilchez S, Hudgens MG, Rogawski ET, Carroll IM, Zambrana LE, Espinoza F, Azcarate-peril MA. Gut Microbiome Composition in Young Nicaraguan Children during Diarrhea Episodes and Recovery. Am J Trop Med Hyg. 2015;93:1187-93. doi:10.4269/ajtmh.15-0322 PMID:26350452

[20] Thompson CN, Phan MVT, Hoang NVM, Minh P V, Vinh NT, Thuy CT, Nga TTT, Rabaa M a., Duy PT, Dung TTN, et al. A Prospective Multi-Center Observational Study of Children Hospitalized with Diarrhea in Ho Chi Minh City, Vietnam. Am J Trop Med Hyg [Internet]. 2015;92:1045-52. Available from: http://www.ajtmh.org/cgi/doi/10.4269/ajtmh.14-0655. doi:10.4269/ajtmh.14-0655

[21] Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE. Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci. 2011;108:4578-85. doi:10.1073/pnas.1000081107 PMID:20668239

[22] Avershina E, Storrø O, Øien T, Johnsen R, Wilson R, Egeland T, Rudia K. Bifidobacterial succession and correlation networks in a large unselected cohort of mothers and their children. Appl Environ Microbiol. 2013;79:497-507. doi:10.1128/AEM.02359-12 PMID:23124244

[23] Singh P, Teal TK, Marsh TL, Tiedje JM, Mosci R, Jernigan K, Zell A, Newton DW, Salimnia H, Lephart P, et al. Intestinal microbial communities associated with acute enteric infections and disease recovery. Microbiome [Internet]. 2015;3:45. Available from: http://www.microbiomejournal.com/content/3/1/45. doi:10.1186/s40168-015-0109-2

[24] Sarker SA, Sultana S, Reuteler G, Moine D, Descombes P, Charton F, Bourdin G, McCallin S, Ngom-Bru C, Neville T, et al. Oral Phage Therapy of Acute Bacterial Diarrhea With Two Coliphage Preparations: A Randomized Trial in Children From Bangladesh. EBioMedicine [Internet]. 2016;4:124-37. Available from: http://www.sciencedirect.com/science/article/pii/S2352396415302516. doi:10.1016/j.ebiom.2015.12.023

[25] Monira S, Hoq MM, Chowdhury AKA, Suau A, Magne F, Endtz Hp, Alam M, Rahman M, Pochart P, Desjeux J-F, et al. Short-chain fatty acids and commensal microbiota in the faeces of severely malnourished children with cholera rehydrated with three different carbohydrates. Eur J Clin Nutr [Internet]. 2010;64:1116-24. Available from: doi:10.1038/ejcn.2010.123

[26] Bradshaw DJ, Marsh PD, Keith Watson G, Allison C. Role of Fusobacterium nucleatum and coaggregation in anaerobe survival in planktonic and biofilm oral microbial communities during aeration. Infect Immun. 1998;66:4729-32. PMID:9746571

[27] Levesque C, Lamothe J, Frenette M. Coaggregation of Streptococcus salivarius with periodontopathogens: evidence for

Downloaded by [113.161.86.225] at 21:47 24 August 2017

involvement of fimbriae in the interaction with Prevotella intermedia. Oral Microbiol Immunol [Internet]. 2003;18:333-7. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12930529. doi:10.1034/j.1399-302X.2003.00085.x

[28] Robrish SA, Oliver C, Thompson J. Sugar metabolism by fusobacteria: regulation of transport, phosphorylation, and polymer formation by Fusobacterium mortiferum ATCC 25557. Infect Immun [Internet]. 1991;59:4547-54. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=259076&tool=pmcentrez&rendertype=abstract.

[29] Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, Ludwig W, Whitman WB. Volume Four The Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes [Internet]. In: Brown DR, Hedlund BP, Paster BJ, Ward NL, Ludwig W, Whitman WB, editors. Bergey's Manual of Systematic Bacteriology. New York: Springer; 2010. page 949.Available from: http://link.springer.com/10.1007/978-0-387-68572-4.

[30] Gupta RS, Bhandari V. Phylogeny and molecular signatures for the phylum Fusobacteria and its distinct subclades. Anaerobe [Internet]. 2014;28:182-98. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21503713. doi:10.1016/j.anaerobe.2014.06.007

[31] Portrait V, Cottenceau G, Pons AM. A Fusobacterium mortiferum strain produces a bacteriocin-like substance (s) inhibiting Salmonella enteritidis. Lett Appl Microbiol [Internet]. 2000;31:115-7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10972711. doi:10.1046/j.1365-2672.2000.00774.x

[32] Traa BS, Fischer Walker CL, Munos M, Black RE. Antibiotics for the treatment of dysentery in children. Int J Epidemiol. 2010;39:70-4. doi:10.1093/ije/dyq024

[33] Moratalla A, Gómez-Hurtado I, Santacruz A, Moya Á, Peiró G, Zapater P, González-Navajas JM, Giménez P, Such J, Sanz Y, et al. Protective effect of Bifidobacterium pseudocatenulatum CECT7765 against induced bacterial antigen translocation in experimental cirrhosis. Liver Int. 2014;34:850-8. doi:10.1111/liv.12380 PMID:24267920

[34] Moratalla A, Gómez-Hurtado I, Moya-Pérez Á, Zapater P, Peiró G, González-Navajas JM, Gómez Del Pulgar EM, Such J, Sanz Y, Francés R. Bifidobacterium pseudocatenulatum CECT7765 promotes a TLR2-dependent anti-inflammatory response in intestinal lymphocytes from mice with cirrhosis. Eur J Nutr [Internet]. 2016;55:197-206. Available from: doi:10.1007/s00394-015-0837-x

[35] Moya-Pérez A, Neef A, Sanz Y. Bifidobacterium pseudocatenulatum CECT 7765 Reduces Obesity-Associated Inflammation by Restoring the Lymphocyte-Macrophage Balance and Gut Microbiota Structure in High-Fat Diet-Fed Mice. PLoS One [Internet]. 2015;10:e0126976. Available from: doi:10.1371/journal.pone.0126976

[36] Moratalla A, Caparrós E, Juanola O, Portune K, Puig-Kröger A, Estrada-Capetillo L, Bellot P, Gómez-Hurtado I, Piñero P, Zapater P, et al. Bifidobacterium pseudocatenulatum CECT7765 induces an M2 anti-inflammatory transition in macrophages from patients with cirrhosis. J Hepatol [Internet]. 2015;64:135-45. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0168827815005929. doi:10.1016/j.jhep.2015.08.020

[37] Asahara T, Shimizu K, Nomoto K, Hamabata T, Ozawa A, Takeda Y. Probiotic Bifidobacteria Protect Mice from Lethal Infection with Shiga Toxin-Producing Escherichia coli O157 : H7. Infect Immun [Internet]. 2004;72:2240-7. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC375161/. doi:10.1128/IAI.72.4.2240-2247.2004

[38] Lopez CA, Kingsbury DD, Velazquez EM, Baumler AJ. Collateral damage: Microbiota-derived metabolites and immune function in the antibiotic era. Cell Host Microbe. 2014;16:156-63. doi:10.1016/j.chom.2014.07.009 PMID:25121745

[39] Duong VT, Phat VV, Tuyen HT, Dung TTN, Trung PD, Minh P Van, Phuong Tu LT, Campbell JI, Phuc H Le, Thanh Ha TT, et al. Evaluation of luminex xTAG gastrointestinal pathogen panel assay for detection of multiple diarrheal pathogens in fecal samples in Vietnam. J Clin Microbiol. 2016;54:1094-100. doi:10.1128/JCM.03321-15 PMID:26865681

[40] Velly H, Britton RA, Preidis GA. Mechanisms of cross-talk between the diet, the intestinal microbiome, and the undernourished host. Gut Microbes [Internet]. 2017;8:98-112. Available from: doi:10.1080/19490976.2016.1267888

[41] Weiss S, Treuren W Van, Lozupone C, Faust K, Friedman J, Deng Y, Xia LC, Xu ZZ, Ursell L. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. ISME J [Internet]. 2016;10:1669-81. Available from: doi:10.1038/ismej.2015.235

[42] World Health Organization. WHO child growth standards: Methods and Development [Internet]. Geneva, Switzerland; 2006. Available from: http://hpps.kbsplit.hr/hpps-2008/pdf/dok03.pdf.

[43] My PVT, Thompson CN, Phuc H Le, Tuyet PTN, Vinh H, Nguyen VMH, Minh P Van, Vinh NT, Thuy CT, Tran TTN, et al. Endemic norovirus infections in children, Ho Chi Minh City, Vietnam, 2009–2010. Emerg Infect Dis [Internet]. 2013;19:29-32. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3713809&tool=pmcentrez&rendertype=abstract%5Cnhttp://europepmc.org/abstract/MED/23735160.

[44] My PVT, Rabaa MA, Vinh H, Holmes EC, Hoang NVM, Vinh NT, Phuong LT, Tham NT, Bay PVB, Campbell JI, et al. The emergence of rotavirus G12 and the prevalence of enteric viruses in hospitalized pediatric diarrheal patients in Southern Vietnam. Am J Trop Med Hyg. 2011;85:768-75. doi:10.4269/ajtmh.2011.11-0364 PMID:21976585

Downloaded by [113.161.86.225] at 21:47 24 August 2017

[45] Ong SH, Kukkillaya VU, Wilm A, Lay C, Ho EXP, Low L, Hibberd ML, Nagarajan N. Species Identification and Profiling of Complex Microbial Communities Using Shotgun Illumina Sequencing of 16S rRNA Amplicon Sequences. PLoS One 2013;8:1-8. doi:10.1371/journal.pone.0060811

[46] Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. Genome Biol [Internet]. 2011;12:R44. Available from: http://genomebiology.com/2011/12/5/R44. doi:10.1186/gb-2011-12-5-r44

[47] Joshi N, Fass J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle. 2011.

[48] Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol [Internet]. 2009 [cited 2014 Jul 9];75:7537-41. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2786419&tool=pmcentrez&rendertype=abstract. doi:10.1128/AEM.01541-09

[49] Edgar RC. UPARSE, highly accurate OTU sequences from microbial amplicon reads. Nat Methods [Internet] 2013 [cited 2014 Jul 10];10:996-8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23955772. doi:10.1038/nmeth.2604

[50] Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS One [Internet]. 2010;5:e9490. Available from: doi:10.1371/journal.pone.0009490

[51] Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol [Internet]. 2007 [cited 2014 Jul 9];73:5261-7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1950982&tool=pmcentrez&rendertype=abstract. doi:10.1128/AEM.00062-07

[52] R Core Team. R: A language and environment for statistical computing. [Internet]. 2016; Available from: https://www.r-project.org/.

[53] McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One [Internet]. 2013 [cited 2014 Jul 14];8:e61217. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3632530&tool=pmcentrez&rendertype=abstract. doi:10.1371/journal.pone.0061217

[54] Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. cluster: Cluster Analysis Basics and Extensions. R package version 2.0.4. 2016.

[55] Ishwaran H, Kogalur UB. Random Forests for Survival, Regression and Classification (RF-SRC), R package version 2.2.0. 2016.

[56] Wickham H. ggplot2 [Internet]. 2009. Available from: http://link.springer.com/10.1007/978-0-387-98141–3%5Cnpapers3://publication/doi/10.1007/978-0-387-98141-3.

[57] Venables W, Ripley B. Modern Applied Statistics with S. Fourth Edi. New York: Springer; 2002.

[58] Zeileis A, Hothorn T. Diagnostic Checking in Regression Relationships. R News 2002;2:7-10.

[59] Oksanen J, Blanchet G, Kindt R, Legendre P, Minchin P, O'Hara R, Simpson G, Solymos P, Stevens M, Wagner H. vegan: Community Ecology Package. R package version 2.3-5. 2016.

[60] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol [Internet]. 2014;15:550. Available from: http://genomebiology.com/2014/15/12/550. doi:10.1186/s13059-014-0550-8

[61] Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and Compositionally Robust Inference of Microbial Ecological Networks. PLOS Comput Biol [Internet]. 2015;11:e1004226. Available from: doi:10.1371/journal.pcbi.1004226

[62] DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, Sun CL, Goltsman DS a, Wong RJ, Shaw G, et al. Temporal and spatial variation of the human microbiota during pregnancy. Proc Natl Acad Sci [Internet]. 2015;112:201502875. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4568272&tool=pmcentrez&rendertype=abstract.

[63] Koren O, Knights D, Gonzalez A, Waldron L, Segata N, Knight R, Huttenhower C, Ley RE. A Guide to Enterotypes across the Human Body: Meta-Analysis of Microbial Community Structures in Human Microbiome Datasets. PLoS Comput Biol [Internet]. 2013;9:e1002863. Available from: doi:10.1371/journal.pcbi.1002863

[64] McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput Biol [Internet]. 2014 [cited 2014 Jul 15];10:e1003531. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3974642&tool=pmcentrez&rendertype=abstract. doi:10.1371/journal.pcbi.1003531

[65] Thorsen J, Brejnrod A, Mortensen M, Rasmussen MA, Stokholm J, Al-Soud WA, Sørensen S, Bisgaard H, Waage J. Large-scale benchmarking reveals false discoveries and count transformation sensitivity in 16S rRNA gene amplicon data analysis methods used in microbiome studies. Microbiome [Internet]. 2016;4:62. Available from: http://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-016-0208-8. doi:10.1186/s40168-016-0208-8

[66] Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, Lozupone C, Zaneveld JR, Vázquez-Baeza Y, Birmingham A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome [Internet]. 2017;5:27. Available from: doi:10.1186/s40168-017-0237-y

[67] Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative beta-diversity measures lead to different insights into factors that structure microbial communities. Appl Environ Microbiol. 2007;73:1576-85. doi:10.1128/AEM.01996-06 PMID:17220268

[68] Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. J R Stat Soc [Internet]. 2001;63:411-23. Available from: doi:10.1111/1467-9868.00293

[69] Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, et al. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci [Internet]. 2011;108:4680-7. Available from: doi:10.1073/pnas.1002611107

[70] MacIntyre DA, Chandiramani M, Lee YS, Kindinger L, Smith A, Angelopoulos N, Lehne B, Arulkumaran S, Brown R, Teoh TG, et al. The vaginal microbiome during pregnancy and the postpartum period in a European population. Sci Rep [Internet]. 2015;5:8988. Available from: http://www.nature.com/srep/2015/150311/srep08988/full/srep08988.html. doi:10.1038/srep08988

[71] Breiman L. Random forests. Mach Learn [Internet] 2001; 45:5-32. Available from: http://link.springer.com/article/10.1023/A:1010933404324. doi:10.1023/A:1010933404324

[72] Egendre PIL, Anderson MJ. Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. Ecol Monogr. 1999;69:1-24. doi:10.1890/0012-9615(1999)069%5b0001:DBRATM%5d2.0.CO;2

[73] Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLoS Comput Biol [Internet]. 2012;8:e1002687. Available from: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.100268. doi:10.1371/journal.pcbi.1002687

Downloaded by [113.161.86.225] at 21:47 24 August 2017

MICROBIAL GENOMICS

MINI REVIEW

Chung The and Baker, *Microbial Genomics* 2018;4

DOI 10.1099/mgen.0.000171

Out of Asia: the independent rise and global spread of fluoroquinolone-resistant *Shigella*

Hao Chung The* and Stephen Baker

Abstract

Shigella are ranked among the most prevalent aetiologies of diarrhoeal disease worldwide, disproportionately affecting young children in developing countries and high-risk communities in developed settings. Antimicrobial treatment, most commonly with fluoroquinolones, is currently recommended for *Shigella* infections to alleviate symptoms and control disease transmission. Resistance to fluoroquinolones has emerged in differing *Shigella* species (*S. dysenteriae*, *flexneri* and *sonnei*) since the turn of the 21st century, originating in endemic areas, and latterly spreading into non-endemic regions. Despite occurring independently, the emergence of fluoroquinolone resistance in these different *Shigella* species shares striking similarities regarding their epidemiology and resistance mechanisms. Here, we review and discuss the current epidemiology of fluoroquinolone-resistant *Shigella* species, particularly in the light of recent genomic insights.

INTRODUCTION

Shigella, a pathogenic genus within the extensive Gram-negative family *Enterobacteriaceae*, is a major cause of diarrhoeal disease worldwide [1, 2]. The global burden of shigellosis is estimated to be 125 million cases per year, of which 160 000 lead to death [2, 3]. The disease disproportionately affects young children in low-income tropical settings, where malnutrition, inadequate sanitation and limited access to clean water appear to facilitate the transmission of the infecting organisms. The genus *Shigella* does not comprise a monophyletic group of organisms but is formed of multiple discrete *Escherichia coli* lineages, all of which harbour a signature virulence plasmid responsible for the distinctive invasive pathogenesis [4, 5]. Current serology classifies the genus into four species or serogroups (*S. dysenteriae*, *boydii*, *flexneri* and *sonnei*), which differ significantly in their epidemiology. Toxigenic *S. dysenteriae* serotype 1 (Sd1) is the causative agent of the now rare, often fatal, epidemic bacillary dysentery. *S. boydii* is only sporadically isolated from diarrhoeal cases in the Indian subcontinent [6–8]. The overwhelming majority of shigellosis cases are presently attributed to *S. flexneri* and *S. sonnei*, which predominantly circulate in developing and developed regions, respectively [9].

Shigellosis usually results in profuse diarrhoea, often accompanied by mucous or bloody discharge. This clinical presentation is associated with disruption of the intestinal epithelium, which is mediated by intracellular proliferation of the infecting *Shigella*. Although the disease is self-limiting, antimicrobial treatment is recommended to prevent further complications, assist recovery and restrict faecal shedding [10, 11]. One of the most commonly prescribed groups of antimicrobials for shigellosis is the fluoroquinolones, which directly interact with the bacterial DNA gyrase (encoded by *gyrA* and *gyrB*) and topoisomerase IV (encoded by *parC* and *parE*) to inhibit functional replication and induce bacterial cell death [12]. Routine surveillance has documented dramatic increases in the frequency of fluoroquinolone-resistant (FQR) *Shigella*, estimating that resistance increased from 0.6 % in 1998–2000 to 29 % in 2007–2009 of the endemic shigellosis in Asia and Africa [9, 13] (Fig. 1). The genetic mechanism(s) underlying resistance is commonly attributed to mutations in the quinolone resistance determining region (QRDR), ultimately diminishing the interaction between the antimicrobial and its target proteins [14]. Resistance to fluoroquinolones narrows ever-dwindling treatment options, placing those who are vulnerable at the increased risk of complications and hampering the efficient management of outbreaks. These factors have placed FQR *Shigella* on the list of global priority pathogens that urgently need focused development of novel antimicrobials [15]. This review aims to summarize the epidemiology of various FQR *Shigella* species, highlighting insights provided through genome sequencing and

Received 23 October 2017; Accepted 20 March 2018

Author affiliation: Enteric Infections, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam.

***Correspondence:** Hao Chung The, haoct@oucru.org

Keywords: *Shigella*; fluoroquinolone resistance; Asia; quinolone resistance determining region (QRDR); epidemiology; genomic.

Abbreviations: FQR, fluoroquinolone-resistant; PFGE, pulse field gel electrophoresis; QRDR, quinolone resistance determining region; Sd1, *Shigella dysenteriae* serotype 1; MSM, men who have sex with men.

000171 © 2018 The Authors

This is an open access article under the terms of the http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Downloaded from www.microbiologyresearch.org by
IP: 14.169.234.47
On: Fri, 06 Apr 2018 11:03:46

phylogenetic reconstruction. Due to its low prevalence and research focus, *S. boydii* will be excluded from this discussion.

Shigella dysenteriae

Resistance against fluoroquinolones had not been previously observed in Sd1 until FQR organisms were isolated during a dysentery outbreak arising in India and Bangladesh in 2002–2003 [16, 17] (Fig. 2). Molecular characterization by pulse field gel electrophoresis (PFGE) revealed that all contemporaneous FQR Sd1 isolates, causing either the outbreak or sporadic episodes across South Asia, belonged to a single clone [18–20]. However, fluoroquinolone resistance was attributed to two different QRDR mutation profiles: *gyrA*-S83L/D87G and S83L/D87N, which were associated with different geographical distributions [19]. These data suggested that the clone may have first acquired a *gyrA*-S83L mutation as early as 1994, later diverging into two FQR subpopulations, which were then characterized by differing secondary mutations. Indeed, a genomic investigation of the global phylogeny of Sd1 concluded that resistance to fluoroquinolones was acquired only once during the species' evolutionary history, conferred by the co-occurrence of *parC*-S80I, *gyrA*-S83L and a secondary *gyrA*-D87 mutation between 1995 and 2002 [21]. This FQR clone belonged to the internationally successful lineage IV, which has witnessed at least nine independent single QRDR mutational events since the 1970s. The emergence of FQR Sd1 was followed by an abrupt decline after the outbreak, hampering routine monitoring and making the prospect of future FQR Sd1 outbreaks unpredictable [20, 22]. Information regarding resistance in alternative *S. dysenteriae* serotypes is limited, probably due to their low prevalence, even in regions where the disease was once highly endemic.

IMPACT STATEMENT

Shigella is a genus of human-adapted bacterial pathogens that cause dysenteric diarrhoea (shigellosis) in developing and developed countries. A specific class of antimicrobials, known as the fluoroquinolones, is recommended for the treatment of shigellosis, but resistance to this group of antimicrobials is rising rapidly within this genus. Here, we have combined available epidemiological and high-resolution genomic data to outline common themes that define the emergence and circulation of fluoroquinolone-resistant *Shigella* species. The information gathered in this review will be useful for determining optimal shigellosis treatment regimens and to tailor public health measures for alerting, containing and preventing the future spread of these antimicrobial-resistant enteric pathogens.

Shigella flexneri

The majority of epidemiological research on *S. flexneri* has been conducted in South Asia and China, where the pathogen's burden remains significant. The first incidences of FQR in *S. flexneri* were documented in eastern and northern China in 2001–2002 [23, 24], and a detailed genetic screen revealed that the majority of these organisms possessed

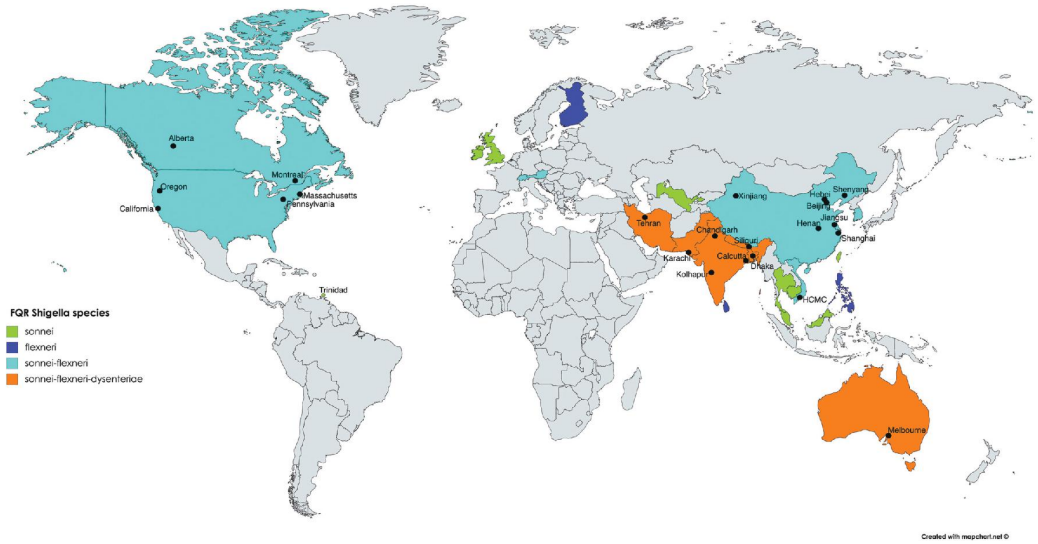


Fig. 1. Worldwide distribution of FQR *Shigella*. Countries are coloured where different FQR *Shigella* species have been reported in the literature (see key). Black filled circles indicate specific regions where FQR *Shigella* have been isolated. Countries with no information or have not reported isolation of FQR *Shigella* are coloured grey. HCMC, Ho Chi Minh city.

Downloaded from www.microbiologyresearch.org by
IP: 14.1[illegible]9.234.47
On: Fri, 06 Apr 2018 11:03:46

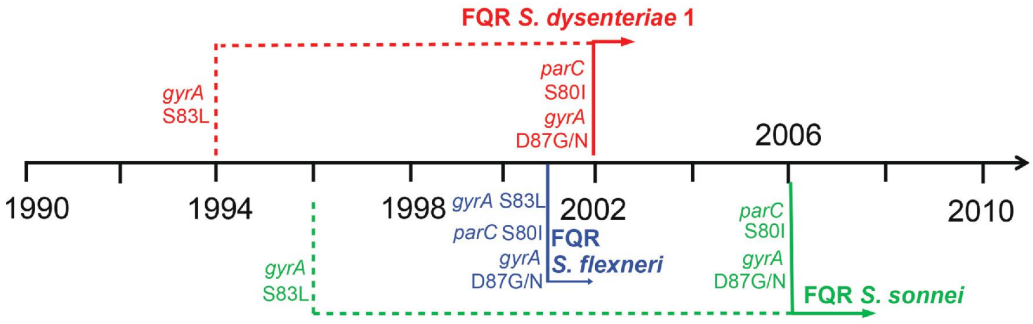


Fig. 2. Timeline detailing the emergences of FQR *Shigella* species. The dashed lines represent the first occurrences of the initial QRDR mutation in the FQR clone if known, as described by epidemiological or genomic data. The solid lines indicate the first reports of FQR *Shigella* species as well as the QRDR mutations that became incorporated into these clones by this designated time. The presumed order of occurrence for these mutations is from top to bottom.

gyrA-S83L, *gyrA*-D87G and *parC*-S80I QRDR mutations [22]. Subsequently, a thorough examination of >2000 Bangladeshi *S. flexneri* underlined a worrying rising trend of fluoroquinolone resistance, which was almost exclusively found in serotype 2a [22]. Although it initially appeared in 2005, fluoroquinolone resistance escalated rapidly and its prevalence was >40 % of all native *S. flexneri* by 2010. These Bangladeshi isolates differed from their Chinese counterparts by a secondary QRDR mutation, harbouring *gyrA*-D87N instead of G. Furthermore, FQR *S. flexneri* with identical mutation profiles were recovered during a decade-long surveillance in Switzerland, highlighting the occurrence of fluoroquinolone resistance in non-endemic regions [25].

Routine dysentery surveillance in China has reported a steady increase of FQR *S. flexneri* of various serotypes, including 1a, 1c, 2a, 2b, 2av, 4a, 4c and X [26–31]. This observation suggests that the FQR phenotype has either emerged on numerous independent occasions across several serotypes or was acquired once, prior to subsequent intensive serotype switching events. Previous genomic studies reported that serotype conversion within a lineage is a commonly observed phenomenon for *S. flexneri* [32, 33]. Available literature provides greater support for the role of serotype switching on creating multiple FQR *S. flexneri* serotypes. Despite being present in a wide range of serotypes and locations, identical QRDR mutations have been frequently encountered in *S. flexneri* in China, encompassing *gyrA*-S83L, *gyrA*-D87G/N and *parC*-S80I. Furthermore, these mutations are commonly accompanied by an unusual mutation (*gyrA*-H211Y), which was also present in the aforementioned Bangladeshi FQR isolates [22, 28, 31, 34]. FQR *S. flexneri* from these two countries were also found to exhibit a close genetic relationship via PFGE [22]. This combined evidence indicates that spatially dispersed FQR *S. flexneri* probably belong to one dominant widespread clone, where the *gyrA*-H211Y, *gyrA*-S83L and *parC*-S80I mutations arose prior to geographical divergence. Later, a secondary mutation in *gyrA* delineated the Bangladeshi (*gyrA*-D87N) and the Chinese (*gyrA*-D87G) FQR isolates. However, the increasing isolation frequency of the *gyrA*-D87N variant in parts of China may be the result of a higher degree of trans-border dissemination from South Asia and/or the separate, indigenous emergence of a competent FQR subclone [31]. Due to the degree of genetic diversity and the complex population structures within *S. flexneri*, the true nature of such events cannot be easily measured using low-resolution molecular typing methods.

Shigella sonnei

A shift in species dominance (from *S. flexneri* to *S. sonnei*) has been observed concurrently in multiple Asian countries as they undergo rapid economic transition; this has been recorded in Bangladesh, China, Thailand and Vietnam [35–39]. This intriguing trend greatly increases the burden of *S. sonnei* worldwide, making antimicrobial resistance in this species a focal target for monitoring. Surveillance studies in developed countries have identified strong epidemiological links between FQR *S. sonnei* and a travel history to India [40, 41]. Moreover, despite disparate spatial distributions, these isolates share the same pulsotype (via PFGE) with FQR *S. sonnei* recovered in South Asia [36, 42–44]. These results suggest that contemporaneous FQR *S. sonnei* are clonal and have evolved and spread in the region before disseminating intercontinentally. Indeed, phylogenetic analysis on representative extant FQR *S. sonnei* confirmed this hypothesis, concluding that South Asia was the most likely origin of these organisms [45]. Furthermore, this study identified two distinct regional diversifications of the FQR clone out of South Asia, with one circulating in Southeast Asia and another appearing to instigate sustained transmission within Europe and America. These observations concur with frequent reports of native FQR *S. sonnei* circulating in Cambodia, Vietnam and California [46–48]. Fluoroquinolone resistance in *S. sonnei* is generally determined by the sequential accumulation of three mutations: *gyrA*-S83L, *parC*-S80I and *gyrA*-D87G [45, 49]. However, other

Downloaded from www.microbiologyresearch.org by
IP: 14.139.234.47
On: Fri, 06 Apr 2018 11:03:46

resistance mechanisms, including differing mutations in QRDR (*gyrA*-D87N instead of D87G) and the synergy between the plasmid-mediated *qnrB* gene and *gyrA* mutations, have also been identified [50, 51]. It is of particular concern that the transmission of the FQR clone is intensified in high-risk contact networks, such as those reported among MSM (men who have sex with men) communities in non-endemic Canada and Taiwan [52, 53]. Therefore, the propagation of FQR *Shigella* should be closely monitored in MSM networks, especially in the wake of increased shigellosis incidence, HIV infection and resistance to other antimicrobials, such as azithromycin, within this high-risk group [54].

OUTLOOK

The presented evidence reveals striking similarities between the emergence of fluoroquinolone resistance among the discrete *Shigella* species. (1) FQR is almost exclusively determined by sequential QRDR mutations in the following order: *gyrA*-S83L, *parC*-S80I and *gyrA*-D87G/N. (2) To date, the majority of FQR isolates identified within an individual species are clonal despite their wide geographical distribution. (3) South Asia, and potentially China, serve as likely reservoirs for the rise and spread of resistant clones. These interpretations are currently deduced from genomic insights into Sd1 and *S. sonnei*, and are subjected to various confounders, including geographical bias in sample collection. However, the extensive genetic diversity within *S. flexneri* may present an alternative scenario, which will benefit from large-scale molecular epidemiology data generated through whole-genome sequencing.

The first widely used fluoroquinolone, ciprofloxacin, was introduced to clinical practice in 1987. However, resistance in *Shigella* only began to emerge in the early 2000s. The intervening period witnessed the emergence of *Shigella* exhibiting resistance to multiple antimicrobials including co-trimoxazole, ampicillin and nalidixic acid [55]. Therefore, fluoroquinolones, such as ciprofloxacin, began to be deployed more commonly to manage drug-resistant shigellosis, and its use became routine, as recommended by the World Health Organization in 2005 [11, 56]. Recent experimental and modelling work into the evolution of fluoroquinolone resistance could offer explanations for the observed pattern between the various *Shigella* species. The ordered QRDR mutations are selected in favour of those in efflux regulatory machinery due to their co-optimization for non-susceptibility (MIC levels) and fitness cost [57, 58]. For both *in vitro* generated and clinical isolates, resistance to fluoroquinolones almost exclusively commences with an initial mutation, *gyrA*-S83L. This mutation has arisen independently on multiple occasions for different *Shigella* lineages, possibly as an adaptive strategy for resistance against the first-generation quinolone, nalidixic acid [21, 32, 59]. However, the key determining factor in this stepwise evolution is the subsequent mutation, *parC*-S80I, which occurs much less frequently during evolution but gives the bacterium a significant increase in fluoroquinolone MIC and potentially a non-inferior fitness. Such a mutation is suggested to be favourably selected in abundance of mutation supply, fulfilled either by a large population size or a high mutation rate, when antimicrobial pressure is high [58]. Given that the mutation rate of the bacterium *Shigella* is relatively stable, the first scenario appears to be more plausible [21, 32, 49, 59]. High population densities in South Asia could promote extensive and sustained *Shigella* transmission, resulting in a large bacterial population. Suboptimal public health measures in the region, exemplified by the fact that only 40 % of the Indian population has access to improved sanitation [60], further amplify the transmission cycle of *Shigella*. This expansion has arisen on a backdrop of rapidly increasing fluoroquinolone use for treatment of multiple enteric and febrile diseases since the turn of this century [56, 61]. Indeed, India, with 12.9 billion units, was ranked as the world's largest antimicrobial consumer in 2010 [62]. These contributing factors might render South Asia a unique focal point for the emergence of human-restricted FQR enteric bacteria, including *Shigella* and *Salmonella* Typhi [63].

CONCLUSION

The emergence of FQR *Shigella* has been quickly followed by the expansion and, for *S. sonnei*, rapid international spread. Furthermore, co-resistance to other first-line antimicrobials, such as the macrolides and third-generation cephalosporins, is frequently identified among these bacteria [47, 50]. These antimicrobial resistance combinations present a serious public health threat for the effective treatment and management of shigellosis. It has been experimentally demonstrated that the described QRDR mutations may carry no detrimental or limited fitness cost to the resistant *Enterobacteriaceae*, even in the absence of fluoroquinolone pressure [57, 64]. Alternatively, fluoroquinolone resistance has been coupled with the successful clonal propagation of several multi-drug-resistant pathogens, including *Staphylococcus aureus*, *Klebsiella pneumoniae*, *E. coli*, *Clostridium difficile* and *Neisseria gonorrhoeae* [65, 66]. All these major FQR clones were found to harbour specific QRDR mutation combinations, indicating that these resistance genotypes induce a minimal fitness disadvantage. Although little is known about the impact of alleviating fluoroquinolone pressure on the clonal dominance of FQR bacteria in nature, we speculate that withdrawal of such pressure in clinical settings is unlikely to discontinue the dominance of FQR *Shigella* in the transmission chain. However, future research is warranted to challenge this hypothesis, as well as to develop the best practices for controlling and treating new emerging antimicrobial-resistant clones.

Funding information
HCT received a DPhil scholarship from the Tropical Network Fund, Nuffield Department of Medicine, University of Oxford. SB is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (100087/Z/12/Z).

Downloaded from www.microbiologyresearch.org by
IP: 14.1[illegible].234.47
On: Fri, 06 Apr 2018 11:03:46

Conflicts of interest

The authors declare that there are no conflicts of interest.

Ethical statement

No ethical approval was required for the research in this study.

References

1. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH *et al.* Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* 2013;382:209–222.
2. Troeger C, Forouzanfar M, Rao PC, Khalil I, Brown A *et al.* Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Infect Dis* 2017;17:909–948.
3. Bardhan P, Faruque AS, Naheed A, Sack DA. Decrease in shigellosis-related deaths without *Shigella* spp.-specific interventions, Asia. *Emerg Infect Dis* 2010;16:1718–1723.
4. Sahl JW, Morris CR, Emberger J, Fraser CM, Ochieng JB *et al.* Defining the phylogenomics of *Shigella* species: a pathway to diagnostics. *J Clin Microbiol* 2015;53:951–960.
5. Chung The H, Thanh DP, Holt KE, Thomson NR, Baker S. The genomic signatures of *Shigella* evolution, adaptation and geographical spread. *Nat Rev Microbiol* 2016;14:235–250.
6. Shiga K. Ueber den erreger der dysenterie in Japan (vorläufige mitteilung). *Zentralbl Bakteriol Mikrobiol* 1898;23:599–600.
7. Pal SC. Epidemic bacillary dysentery in West Bengal, India, 1984. *The Lancet* 1984;323:1462.
8. Barceloux DG. *Shigella* species (Shiga Enterotoxins). In: *Medical Toxicology of Natural Substances*. Hoboken, New Jersey, USA: John Wiley & Sons, Inc; 2008. pp. 150–155.
9. Gu B, Cao Y, Pan S, Zhuang L, Yu R *et al.* Comparison of the prevalence and changing resistance to nalidixic acid and ciprofloxacin of Shigella between Europe-America and Asia-Africa from 1998 to 2009. *Int J Antimicrob Agents* 2012;40:9–17.
10. Vinh H, Wain J, Chinh MT, Tam CT, Trang PT *et al.* Treatment of bacillary dysentery in Vietnamese children: two doses of ofloxacin versus 5-days nalidixic acid. *Trans R Soc Trop Med Hyg* 2000;94:323–326.
11. WHO. Guidelines for the control of shigellosis, including epidemics due to *Shigella dysenteriae* type 1. World Health Organization. 2005.
12. Redgrave LS, Sutton SB, Webber MA, Piddock LJ. Fluoroquinolone resistance: mechanisms, impact on bacteria, and role in evolutionary success. *Trends Microbiol* 2014;22:438–445.
13. Kahsay AG, Muthupandian S. A review on Sero diversity and antimicrobial resistance patterns of Shigella species in Africa, Asia and South America, 2001–2014. *BMC Res Notes* 2016;9:422.
14. Tamanna, Ramana J. Structural insights into the fluoroquinolone resistance mechanism of *Shigella flexneri* DNA gyrase and topoisomerase IV. *Microb Drug Resist* 2016;22:404–411.
15. Tacconelli E, Magrini N, Kahlmeter G, Singh N. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. *World Heal Organ* 2017:1–7.
16. Bhattacharya SK, Sarkar K, Balakrish Nair G, Faruque AS, Sack DA. Multidrug-resistant *Shigella dysenteriae* type 1 in south Asia. *Lancet Infect Dis* 2003;3:755.
17. Naheed A, Kalluri P, Talukder KA, Faruque AS, Khatun F *et al.* Fluoroquinolone-resistant *Shigella dysenteriae* type 1 in northeastern Bangladesh. *Lancet Infect Dis* 2004;4:607–608.
18. Pazhani GP, Sarkar B, Ramamurthy T, Bhattacharya SK, Takeda Y *et al.* Clonal multidrug-resistant *Shigella dysenteriae* type 1 strains associated with epidemic and sporadic dysenteries in eastern India. *Antimicrob Agents Chemother* 2004;48:681–684.
19. Talukder KA, Khajanchi BK, Islam MA, Dutta DK, Islam Z *et al.* Genetic relatedness of ciprofloxacin-resistant *Shigella dysenteriae* type 1 strains isolated in south Asia. *J Antimicrob Chemother* 2004;54:730–734.
20. Taneja N. Changing epidemiology of shigellosis and emergence of ciprofloxacin-resistant *Shigellae* in India. *J Clin Microbiol* 2007;45:678–679.
21. Njamkepo E, Fawal N, Tran-Dien A, Hawkey J, Strockbine N *et al.* Global phylogeography and evolutionary history of *Shigella dysenteriae* type 1. *Nat Microbiol* 2016;1:16027.
22. Azmi IJ, Khajanchi BK, Akter F, Hasan TN, Shahnaij M *et al.* Fluoroquinolone resistance mechanisms of *Shigella flexneri* isolated in Bangladesh. *PLoS One* 2014;9:e102533.
23. Wang XY, Du L, von Seidlein L, Xu ZY, Zhang YL *et al.* Occurrence of shigellosis in the young and elderly in rural China: results of a 12-month population-based surveillance study. *Am J Trop Med Hyg* 2005;73:416–422.
24. Qin T, Bi R, Fan W, Kang H, Ma P *et al.* Novel mutations in quinolone resistance-determining regions of gyrA, gyrB, parC and parE in *Shigella flexneri* clinical isolates from eastern Chinese populations between 2001 and 2011. *Eur J Clin Microbiol Infect Dis* 2016;35:2037–2045.
25. Nüesch-Inderbinen M, Heini N, Zurfluh K, Althaus D, Hächler H *et al.* *Shigella* antimicrobial drug resistance mechanisms, 2004–2014. *Emerg Infect Dis* 2016;22:1083–1085.
26. Zhang J, Jin H, Hu J, Yuan Z, Shi W *et al.* Antimicrobial resistance of *Shigella* spp. from humans in Shanghai, China, 2004–2011. *Diagn Microbiol Infect Dis* 2014;78:282–286.
27. Qiu S, Xu X, Wang Y, Yang G, Wang Z *et al.* Emergence of resistance to fluoroquinolones and third-generation cephalosporins in *Shigella flexneri* subserotype 1c isolates from China. *Clin Microbiol Infect* 2012;18:E95–E98.
28. Cui X, Wang J, Yang C, Liang B, Ma Q *et al.* Prevalence and antimicrobial resistance of *Shigella flexneri* serotype 2 variant in China. *Front Microbiol* 2015;6:435.
29. Pu XY, Pan JC, Wang HQ, Zhang W, Huang ZC *et al.* Characterization of fluoroquinolone-resistant *Shigella flexneri* in Hangzhou area of China. *J Antimicrob Chemother* 2009;63:917–920.
30. Zhang W, Luo Y, Li J, Lin L, Ma Y *et al.* Wide dissemination of multidrug-resistant *Shigella* isolates in China. *J Antimicrob Chemother* 2011;66:2527–2535.
31. Qin T, Qian H, Fan W, Ma P, Zhou L *et al.* Newest data on fluoroquinolone resistance mechanism of *Shigella flexneri* isolates in Jiangsu Province of China. *Antimicrob Resist Infect Control* 2017;6:1–8.
32. Connor TR, Barker CR, Baker KS, Weill FX, Talukder KA *et al.* Species-wide whole genome sequencing reveals historical global spread and recent local persistence in *Shigella flexneri*. *Elife* 2015;4:1–16.
33. Zhang N, Lan R, Sun Q, Wang J, Wang Y *et al.* Genomic portrait of the evolution and epidemic spread of a recently emerged multidrug-resistant *Shigella flexneri* clone in China. *J Clin Microbiol* 2014;52:1119–1126.
34. Yang C, Li P, Zhang X, Ma Q, Cui X *et al.* Molecular characterization and analysis of high-level multidrug-resistance of *Shigella flexneri* serotype 4s strains from China. *Sci Rep* 2016;6:29124.
35. Thompson CN, Duy PT, Baker S. The rising dominance of *Shigella sonnei*: an intercontinental shift in the etiology of bacillary dysentery. *PLoS Negl Trop Dis* 2015;9:e0003708.
36. Ud-Din AI, Wahid SU, Latif HA, Shahnaij M, Akter M *et al.* Changing trends in the prevalence of *Shigella* species: emergence of multi-drug resistant *Shigella sonnei* biotype g in Bangladesh. *PLoS One* 2013;8:e82601.
37. Qiu S, Xu X, Yang C, Wang J, Liang B *et al.* Shift in serotype distribution of *Shigella* species in China, 2003-2013. *Clin Microbiol Infect* 2015;21:252.e5–252.e8.

Downloaded from www.microbiologyresearch.org by
IP: 14.1[illegible]9.234.47
On: Fri, 06 Apr 2018 11:03:46

38. Bangtrakulnonth A, Vieira AR, Lo Fo Wong DM, Pornreongwong S, Pulsrikarn C *et al.* *Shigella* from humans in Thailand during 1993 to 2006: spatial-time trends in species and serotype distribution. *Foodborne Pathog Dis* 2008;5:773–784.
39. Vinh H, Nhu NT, Nga TV, Duy PT, Campbell JI *et al.* A changing picture of shigellosis in southern Vietnam: shifting species dominance, antimicrobial susceptibility and clinical presentation. *BMC Infect Dis* 2009;9:204.
40. De Lappe N, O'Connor J, Garvey P, Mckeown P, Cormican M. Ciprofloxacin-resistant *Shigella sonnei* associated with travel to India. *Emerg Infect Dis* 2015;21:894–896.
41. Bowen A, Hurd J, Hoover C, Khachadourian Y, Traphagen E . *et al.* Importation and domestic transmission of *Shigella sonnei* resistant to ciprofloxacin – United States, May 2014-February 2015. *MMWR Morb Mortal Wkly Rep* 2015;64:318–320.
42. Nandy S, Dutta S, Ghosh S, Ganai A, Rajahamsan J *et al.* Foodborne-associated *Shigella sonnei*, India, 2009 and 2010. *Emerg Infect Dis* 2011;17:2072–2073.
43. Ghosh S, Pazhani GP, Chowdhury G, Guin S, Dutta S *et al.* Genetic characteristics and changing antimicrobial resistance among *Shigella* spp. isolated from hospitalized diarrhoeal patients in Kolkata, India. *J Med Microbiol* 2011;60:1460–1466.
44. Ruekit S, Wangchuk S, Dorji T, Tshering KP, Pootong P *et al.* Molecular characterization and PCR-based replicon typing of multidrug resistant *Shigella sonnei* isolates from an outbreak in Thimphu, Bhutan. *BMC Res Notes* 2014;7:95–99.
45. Chung The H, Rabaa MA, Pham Thanh D, De Lappe N, Cormican M *et al.* South Asia as a reservoir for the global spread of ciprofloxacin-resistant *Shigella sonnei*: a cross-sectional study. *PLoS Med* 2016;13:e1002055.
46. Poramathikul K, Bodhidatta L, Chiek S, Oransathid W, Ruekit S *et al.* Multidrug-resistant *Shigella* infections in patients with diarrhea, Cambodia, 2014–2015. *Emerg Infect Dis* 2016;22:1640–1643.
47. Kim JS, Kim JJ, Kim SJ, Jeon S, Seo KY *et al.* *Shigella sonnei* associated with travel to Vietnam, Republic of Korea. *Emerg Infect Dis* 2015;21:1247–1250.
48. Kozyreva VK, Jospin G, Greninger AL, Watt JP, Eisen JA *et al.* Recent outbreaks of Shigellosis in California caused by two distinct populations of *Shigella sonnei* with either increased virulence or fluoroquinolone resistance. *mSphere* 2016;1:e00344-16.
49. Chung The H, Rabaa MA, Thanh DP, Ruekit S, Wangchuk S *et al.* Introduction and establishment of fluoroquinolone-resistant *Shigella sonnei* into Bhutan. *Microb Genom* 2015;1:1–11.
50. Sadouki Z, Day MR, Doumith M, Chattaway MA, Dallman TJ *et al.* Comparison of phenotypic and WGS-derived antimicrobial resistance profiles of *Shigella sonnei* isolated from cases of diarrhoeal disease in England and Wales, 2015. *J Antimicrob Chemother* 2017;72:2496–2502.
51. Das A, Natarajan M, Mandal J. The emergence of quinolone resistant *Shigella sonnei*, Pondicherry, India. *PLoS One* 2016;11: e0160290.
52. Gaudreau C, Ratnayake R, Pilon PA, Gagnon S, Roger M *et al.* Ciprofloxacin-resistant *Shigella sonnei* among men who have sex with men, Canada, 2010. *Emerg Infect Dis* 2011;17:1747–1750.
53. Chiou CS, Izumiya H, Kawamura M, Liao YS, Su YS *et al.* The worldwide spread of ciprofloxacin-resistant *Shigella sonnei* among HIV-infected men who have sex with men, Taiwan. *Clin Microbiol Infect* 2016;22:383.e11–383.e16.
54. Mohan K, Hibbert M, Rooney G, Canvin M, Childs T *et al.* What is the overlap between HIV and shigellosis epidemics in England: further evidence of MSM transmission? *Sex Transm Infect* 2018; 94:67–71.
55. Niyogi SK. Increasing antimicrobial resistance–an emerging problem in the treatment of shigellosis. *Clin Microbiol Infect* 2007;13: 1141–1143.
56. Kotwani A, Chaudhury RR, Holloway K. Antibiotic-prescribing practices of primary care prescribers for acute diarrhea in New Delhi, India. *Value Health* 2012;15:S116–S119.
57. Marcusson LL, Frimodt-Møller N, Hughes D. Interplay in the selection of fluoroquinolone resistance and bacterial fitness. *PLoS Pathog* 2009;5:e1000541.
58. Huseby DL, Pietsch F, Brandis G, Garoff L, Tegehall A *et al.* Mutation supply and relative fitness shape the genotypes of ciprofloxacin-resistant *Escherichia coli*. *Mol Biol Evol* 2017;34: 1029–1039.
59. Holt KE, Baker S, Weill FX, Holmes EC, Kitchen A *et al.* *Shigella sonnei* genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. *Nat Genet* 2012;44: 1056–1059.
60. World Bank. 2015. World development indicators. Available from: https://data.worldbank.org/indicator/.
61. Laxminarayan R, Chaudhury RR. Antibiotic resistance in India: drivers and opportunities for action. *PLoS Med* 2016;13:e1001974.
62. van Boeckel TP, Gandra S, Ashok A, Caudron Q, Grenfell BT *et al.* Global antibiotic consumption 2000 to 2010: an analysis of national pharmaceutical sales data. *Lancet Infect Dis* 2014;14: 742–750.
63. Wong VK, Baker S, Pickard DJ, Parkhill J, Page AJ *et al.* Phylogeographical analysis of the dominant multidrug-resistant H58 clade of *Salmonella* Typhi identifies inter- and intracontinental transmission events. *Nat Genet* 2015;47:632–639.
64. Baker S, Duy PT, Nga TV, Dung TT, Phat VV *et al.* Fitness benefits in fluoroquinolone-resistant *Salmonella* Typhi in the absence of antimicrobial pressure. *Elife* 2013;2:e01229.
65. Fuzi M, Szabo D, Csercsik R. Double-serine fluoroquinolone resistance mutations advance major international clones and lineages of various multi-drug resistant bacteria. *Front Microbiol* 2017;8:1–14.
66. Grad YH, Harris SR, Kirkcaldy RD, Green AG, Marks DS *et al.* Genomic epidemiology of gonococcal resistance to extended-spectrum cephalosporins, macrolides, and fluoroquinolones in the United States, 2000–2013. *J Infect Dis* 2016;214:1579–1587.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.

Downloaded from www.microbiologyresearch.org by
IP: 14.1[illegible]0.234.47
On: Fri, 06 Apr 2018 11:03:46