Whole genome sequence analysis of *Shigella* from Malawi identifies fluoroquinolone resistance

George E. Stenhouse¹, Khuzwayo C. Jere^{1,2,3}, Chikondi Peno², Rebecca J. Bengtsson¹, End Chinyama², Jonathan Mandolo², Amy K. Cain⁴, Miren Iturriza-Gómara³, Naor Bar-Zeev⁵, Nigel A. Cunliffe^{1,3}, Jennifer Cornick^{^1,2}, Kate S. Baker^{^1,3*}

Author affiliations:

¹ University of Liverpool, Liverpool, UK; ² Malawi-Liverpool-Wellcome Trust Clinical Research Programme, College of Medicine, University of Malawi, Blantyre, Malawi; ³ NIHR Health Protection Research Unit in Gastrointestinal Infections, University of Liverpool, Liverpool, UK; ⁴ ARC Centre of Excellence in Synthetic Biology, Department of Molecular Sciences, Macquarie University, North Ryde, Australia, ⁵ International Vaccine Access Center Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA

[^] These authors contributed equally to this work

* Corresponding author

Figures and Tables

Isolate ID	Contig length (bp)	Resistance genes	Predicted resistance drug class
22204_7#74	9131 ^o	<i>tet(A)</i>	Tetracycline
(<i>S. boydii</i> 2)	2798	<i>aph(3'')-Ib, aph(6)-Id, sul2</i>	Aminoglycoside, sulfonamide
	2353	<i>dfrA7</i>	Trimethoprim
	1831	<i>blaTEM-1B</i>	Aminopenicillin
22204_7#78	48082	<i>tet(A)</i>	Tetracycline
(<i>S. flexneri</i> 4av)	17334	<i>aph(3'')-Ib, aph(6)-Id, sul2, blaTEM-1B, qnrS1</i>	Aminoglycoside, sulfonamide, aminopenicillin, fluoroquinolone
	1588	<i>dfrA14</i>	Trimethoprim
22204_7#79	34138	<i>tet(A), sul1, dfrA7, catA1</i>	Tetracycline, sulfonamide, trimethoprim, chloramphenicol
(<i>S. flexneri</i> 6)	6200	<i>aph(3'')-Ib, aph(6)-Id, sul2</i>	Aminoglycoside, sulfonamide
22204_7#80	25733	<i>tet(A), sul1, dfrA7, catA1</i>	Tetracycline, sulfonamide, trimethoprim, chloramphenicol
(<i>S. flexneri</i> 6)	6200	<i>aph(3'')-Ib, sul2</i>	Aminoglycoside, sulfonamide
22204_7#81	11392 ^r	<i>dfrA14, sul2, tet(A)</i>	Trimethoprim, sulfonamide, tetracycline
(<i>S. flexneri</i> 3a)			
22204_7#82	45377	<i>tet(B)</i>	Tetracycline
(<i>S. flexneri</i> 3a)	8773	<i>aadA1, blaOXA-1, catA1</i>	Aminoglycoside, aminopenicillin, chloramphenicol
	6790 ^r	<i>aph(6)-Id, dfrA14, sul2</i>	Aminoglycoside, trimethoprim, sulfonamide

Table 1. Antimicrobial resistance genotypic and predicted phenotypic profiles of Malawian Shigella isolates by contiguous sequence.

^o StarAMR identified IncFIB(K) plasmid, ^r StarAMR identified MDR IncQ1 plasmid, ^r possible multi-copy plasmid.

Figure 1. Maximum-likelihood phylogeny of eight isolates from Malawi, contextualised among *Escherichia coli* and *Shigella* and highlighting some AMR profiles of interest.

AMR profiles of interest are those of contiguous sequences which are MDR and shared across multiple isolates, and in one isolate carries additional AMR genes including an FQR gene (*qnrSI*). They are indicated by columns to the right of the tree (A) and by coloured bars in AMR profile chart (B).

A. Columns to the right of phylogeny indicate, from left to right, species, study isolate, AMR profiles of interest. GTR+G substitution model, 1000 bootstrap validation and mid-point rooted. * = *E. coli* clade 3, ^ = *S. flexneri* phylogroup 2, ^γ = *S. flexneri* phylogroup 7. All bootstrap values for internal nodes with support <900 are displayed. Scale bar is in SNPs per site. **B.** Intersection of individual AMR genes by isolate and AMR gene profile by contig.

Abstract

Increasing antimicrobial resistance and limited alternative treatments led to fluoroquinolone resistant *Shigella* strain inclusion on the WHO global priority pathogens list. In this study we characterised multiple *Shigella* isolates from Malawi with whole genome sequence analysis, identifying the acquirable fluoroquinolone resistance determinant *qnrS1*.

Introduction

Shigella is the second leading cause of diarrhoeal death globally with the greatest disease burdens seen in low and middle-income countries (1). Approximately one third of these deaths are in children under the age of five, with infection potentially causing chronic health effects (1, 2). Antimicrobial resistance (AMR) increasingly limits treatment options, threatening to reverse hard won reductions in diarrhoeal mortality in high-burden areas (3). Meanwhile, vaccines are still in development. There is, therefore, a need for effective public health solutions.

Increasing prevalence of fluoroquinolone resistant (FQR) strains and limited, widely effective alternatives to this first-line treatment, has led to these strains being included on the WHO global priority pathogens list (4). Resistance can be acquired *de novo* through a double mutation in the *gyrA* gene (amino acids 83 and 87) of the quinolone resistance determining region (QRDR), with a third mutation in the *parC* gene (AA80) ameliorating the fitness cost (5). Resistance can also be acquired through horizontal transmission of FQR genes (5).

Shigella is reported as a leading cause of diarrhoea among hospitalised children in Malawi, however, there is little information on the circulating strains (6). Whole genome sequence analysis (WGS) has been successfully applied to investigate *Shigella* epidemiology and AMR determinants, and can greatly aid in disease control in high-burden areas, such as Malawi (7). Here, we have applied WGS to characterise *Shigella* strains and AMR determinants in Malawi. Providing important baseline information for public health

interventions including antibiotic treatment and deployment of *Shigella* vaccines, the development of which is a WHO priority (8).

The study

All biochemically-confirmed shigellae collected during a rotavirus vaccine evaluation programme between 2012 and 2015, isolated from faecal samples collected from children under the age of 5 hospitalised with acute gastroenteritis at the Queen Elizabeth Central Hospital, Blantyre, Malawi, were subjected to whole genome sequencing (Supplementary (9)). Species were confirmed with a maximum-likelihood phylogeny of the *Shigella*/*Escherichia coli* clade, generated from a core single nucleotide polymorphism (SNP) alignment (40075 SNPs) using quality trimmed reads mapped to a complete reference genome (Supplementary). Only those genomically-confirmed as *Shigella* or *E. coli* (8/10) were included in the study. All *Shigella* isolates were serotyped *in silico* with ShigaTyper (v1.0.6, <https://github.com/CFSAN-Biostatistics/shigatyper>).

From our eight isolates, four *Shigella* serotypes were identified: two *S. flexneri* 3a (Sf3a, Phylogroup 2), one *S. flexneri* 4av (Sf4av, Phylogroup 7), two *S. flexneri* 6 (Sf6) and one *S. boydii* 2 (Sb, Clade 3). Two isolates were *E. coli* (Figure 1A). The lack of *S. sonnei* isolates in our collection was expected as *S. sonnei*, though highly prevalent globally, is typically associated with high-income nations and industrialisation (10). Prevalence of *S. boydii* in Africa is thought to be low, isolated in 5.7% of cases of a multi-national study (11). However, detection with our small sample size would still be likely by chance (binomial test p-value = 0.375) and may not be due to higher prevalence in Malawi. The strain diversity indicates several, distinct strains circulate in Malawi; further characterisation is needed to aid effective vaccine development for the region.

Identification of *E. coli* among the samples is likely due to the close relatedness between the two species. *Shigella* is a specialised pathovar of *E. coli*, sharing a disease

phenotype with enteroinvasive *E. coli* (EIEC), that is highly adapted to humans (12). To determine whether the *E. coli* isolates were EIEC we looked for the presence of the *mxi-spa* locus, found on the large virulence plasmid (pINV). This locus encodes a Type 3 Secretion System and secreted effector proteins which produce the distinctive invasive disease phenotype (12, 13). Contiguous sequences (contigs) from quality-assessed isolate draft-genomes, assembled using Unicycler (v0.4.7) (14), were compared against a reference EIEC pINV to identify locus presence (Supplementary). Using this approach, one *E. coli* isolate was identified as EIEC (99% identity and 100% coverage with *mxi-spa* locus).

As resistance is a growing hinderance to effective treatment, we determined genotypic AMR profiles for each isolate using starAMR (v0.5.1; <https://github.com/phac-nml/staramr>; Supplementary). All *Shigella* isolates were predicted to be multi-drug resistant (MDR), carrying genes conferring resistance to three or more drug classes, demonstrating MDR *Shigella* circulate in Malawi (Table 1). While overall resistance was high, we observed limited diversity in AMR genes and predicted resistance profiles; 13 genes encoded resistance to 7 antimicrobial classes (Figure 1B, Table 1). This suggests that there are treatments which remain effective in Malawi, such as azithromycin, though this would need confirming in a larger study and may change with mass drug administration programs.

Mobile genetic elements (MGE) are important drivers of AMR dissemination, so we explored the genetic context of our AMR genes to look for evidence of them being within MGEs (15). Resistance contigs from two isolates were identified as likely plasmid contigs, also using starAMR (supplementary): one Sf3 isolate carried a MDR IncQ1 plasmid, and the Sf4av isolate carried an AMR IncFIB(K) plasmid (Table 1). Another MDR contig had a read depth 5.29-fold higher than the chromosomal contigs, possible evidence of a multi-copy plasmid (Table 1). Meanwhile, the same AMR gene profile was identified in contigs from multiple isolates across distinct phylogroups (Sf4av and Sf6), possibly indicating an MGE

(Figure 1, Table 1). While there was limited evidence, due to the limitations of short read sequencing, taken together, our data supports a role for MGEs in the spread of MDR *Shigella* in Malawi.

To predict FQR, we looked for the presence of either point mutations in the QRDR, or FQR-associated genes. Point mutations were identified by comparing the *gyrA* and *parC* amino acid sequences across all isolates, with amino acid identity at resistance-associated sites confirmed as expected based on the literature. One isolate (Sf4av) was predicted FQR, due to the presence of *qnrS1* (Table 1). A mutation at *parC* R91Q was also identified, however, the same variation was present in all isolates, *E. coli* and *Shigella*, and likely represents natural variation rather than a resistance adaptation. Unfortunately, phenotype data were unavailable to confirm these findings.

Detection of *qnrS1* indicates acquirable FQR is present among *Shigella* in Malawi. As this gene is typically carried on a plasmid, a BLAST comparison of the FQR contig (17334 bases) against the NCBI nt database was performed; more accurate methods of plasmid detection were unavailable due to the short sequence reads (5). Three hits against *S. flexneri* plasmids (KY848295.1, CP024474.1, LR213454.1) were found (e-score = 0.0, query sequence coverage and identity $\geq 99.9\%$), suggesting that this gene might be carried on a plasmid. This would mean a high risk of widespread FQR in Malawi and neighbouring regions; further study into the prevalence and nature of FQR in the region is needed.

Conclusions

Together, the high proportion of MDR *Shigella*, the acquirable FQR and the AMR plasmids detected in this study show that, without intervention, controlling shigellosis in Malawi will be increasingly difficult. Highlighting the importance of further research into the epidemiology of *Shigella* in the region to ensure effective disease control.

References

1. Khalil IA, Troeger C, Blacker BF, Rao PC, Brown A, Atherly DE, et al. Morbidity and mortality due to shigella and enterotoxigenic Escherichia coli diarrhoea: the Global Burden of Disease Study 1990–2016. *Lancet Infectious Disease* 2018 25 September 2018;18:1229 - 40.
2. Troeger C, Forouzanfar M, Rao PC, Khalil I, Brown AE, Reiner RC, 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 infectious diseases*. 2017 June 1 2017;17:909-48.
3. Gu B, Cao Y, Pan S, Zhuang L, Yu R, Peng Z, 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. *International Journal of Antimicrobial Agents*. 2012 2012/07/01;40(1):9-17.
4. WHO. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. WHO press. 2017:7.
5. Nüesch-Inderbinen M, Heini N, Zurfluh K, Althaus D, Hächler H, Stephan R. Shigella Antimicrobial Drug Resistance Mechanisms, 2004-2014. *Emerging infectious diseases*. 2016;22(6):1083-5.
6. Iturriza-Gomara M, Jere KC, Hungerford D, Bar-Zeev N, Shioda K, Kanjerwa O, et al. Etiology of Diarrhea Among Hospitalized Children in Blantyre, Malawi, Following Rotavirus Vaccine Introduction: A Case-Control Study. *J Infect Dis*. 2019 Jun 19;220(2):213-8.
7. Baker KS, Campos J, Pichel M, Della Gaspera A, Duarte-Martinez F, Campos-Chacon E, et al. Whole genome sequencing of Shigella sonnei through PulseNet Latin

America and Caribbean: advancing global surveillance of foodborne illnesses. Clin Microbiol Infect. 2017 Nov;23(11):845-53.

8. WHO. DRAFT WHO Preferred Product Characteristics for Vaccines against Shigella. WHO press. 2020.

9. Bar-Zeev N, Kapanda L, Tate JE, Jere KC, Iturriza-Gomara M, Nakagomi O, et al. Effectiveness of a monovalent rotavirus vaccine in infants in Malawi after programmatic roll-out: an observational and case-control study. The Lancet Infectious Diseases. 2015;15(4):422-8.

10. Kotloff KL, Riddle MS, Platts-Mills JA, Pavlinac P, Zaidi AKM. Shigellosis. The Lancet. 2018;391(10122):801-12.

11. Livio S, Strockbine NA, Panchalingam S, Tennant SM, Barry EM, Marohn ME, et al. Shigella isolates from the global enteric multicenter study inform vaccine development. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America. 2014;59(7):933-41.

12. Lan R, Lumb B, Ryan D, Reeves PR. Molecular Evolution of Large Virulence Plasmid in Shigella Clones and Enteroinvasive Escherichia coli. Infection and Immunity. 2001;69(10):6303-9.

13. Schroeder GN, Hilbi H. Molecular pathogenesis of Shigella spp.: controlling host cell signalling, invasion, and death by type III secretion. Clin Microbiol Rev. 2008 Jan;21(1):134-56.

14. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS computational biology. 2017 Jun;13(6):e1005595.

15. Baker KS, Dallman TJ, Field N, Childs T, Mitchell H, Day M, et al. Horizontal antimicrobial resistance transfer drives epidemics of multiple *Shigella* species. *Nat Commun*. 2018 Apr 13;9(1):1462.

Acknowledgements

GES is supported by a studentship from the MRC Discovery Medicine North (DiMeN) Doctoral Training Partnership (MR/N013840/1). KCJ is supported by a Wellcome International Training Fellowship (201945/Z/16/Z). KSB is supported by a Wellcome Trust Clinical Research Career Development Fellowship (106690/A/14/Z). AKC was supported by an Australian Research Council (ARC) DECRA fellowship (DE180100929). The authors would like to acknowledge the technical assistance of Trevor Wilson and Chikondi Jassi, and the bioinformatic assistance of Caisey Pulford and Lewis Fisher.

Nigel Cunliffe is affiliated to the National Institute for Health Research (NIHR) Health Protection Research Unit in Gastrointestinal Infections at University of Liverpool, in partnership with Public Health England, in collaboration with University of Warwick. The views expressed are those of the author(s) and not necessarily those of the NIHR, the Department of Health and Social Care or Public Health England.

Ethical approval was obtained from the National Health Sciences Research Committee, Lilongwe, Malawi (#867) and the Research Ethics Committee of the University of Liverpool, Liverpool, UK (000490).

Tree scale: 4008

Species

S. boydii

S. flexneri

S. sonnei

S. dysenteriae

E. coli

Malawi isolate

AMR contig profile of interest

sul1, dfrA7, catA1, tet(A)

aph(3'')-Ib, aph(6)-Id, sul2

aph(3'')-Ib, aph(6)-Id, sul2, blaTEM-1B, qnrS1

bioRxiv preprint doi: <https://doi.org/10.1101/2020.11.20.391235>; this version posted November 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.