



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **Acquisition and loss of CTX-M plasmids in *Shigella* species associated with MSM transmission in the UK**

**Citation for published version:**

Locke, RK, Greig, DR, Jenkins, C, Dallman, TJ & Cowley, LA 2021, 'Acquisition and loss of CTX-M plasmids in *Shigella* species associated with MSM transmission in the UK', *Microbial Genomics*, vol. 7, no. 8. <https://doi.org/10.1099/mgen.0.000644>

**Digital Object Identifier (DOI):**

[10.1099/mgen.0.000644](https://doi.org/10.1099/mgen.0.000644)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Microbial Genomics

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Acquisition and loss of CTX-M plasmids in *Shigella* species associated with MSM transmission in the UK

Rebecca K. Locke<sup>1,\*</sup>, David R. Greig<sup>2,3</sup>, Claire Jenkins<sup>2</sup>, Tim J. Dallman<sup>2,3</sup> and Lauren A. Cowley<sup>1</sup>

## Abstract

Shigellosis in men who have sex with men (MSM) is caused by multidrug resistant *Shigellae*, exhibiting resistance to antimicrobials including azithromycin, ciprofloxacin and more recently the third-generation cephalosporins. We sequenced four *bla*<sub>CTX-M-27</sub>-positive MSM *Shigella* isolates (2018–20) using Oxford Nanopore Technologies; three *S. sonnei* (identified as two MSM clade 2, one MSM clade 5) and one *S. flexneri* 3a, to explore AMR context. All *S. sonnei* isolates harboured Tn7/Int2 chromosomal integrons, whereas *S. flexneri* 3a contained the *Shigella* Resistance Locus. All strains harboured IncFII pKSR100-like plasmids (67–83kbp); where present *bla*<sub>CTX-M-27</sub> was located on these plasmids flanked by IS26 and IS903B, however *bla*<sub>CTX-M-27</sub> was lost in *S. flexneri* 3a during storage between Illumina and Nanopore sequencing. IncFII AMR regions were mosaic and likely reorganised by IS26; three of the four plasmids contained azithromycin-resistance genes *erm(B)* and *mph(A)* and one harboured the pKSR100 integron. Additionally, all *S. sonnei* isolates possessed a large IncB/O/K/Z plasmid, two of which carried *aph(3')-Ib/aph(6)-Id/sul2* and *tet(A)*. Monitoring the transmission of mobile genetic elements with co-located AMR determinants is necessary to inform empirical treatment guidance and clinical management of MSM-associated shigellosis.

## DATA SUMMARY

Whole genomes sequencing projects for strains 598080 and 607387 are deposited in GenBank under the accession numbers JAENSM000000000 and JAEMEC000000000 respectively. For strain 888048, the chromosomal sequence can be retrieved with the accession CP066809 and its plasmids under accessions MW396860-3. For strain 893916, the chromosomal sequence is deposited under accession CP066810 and plasmid sequences under accessions MW396858-9 and MW396864. Further information can be found in Table S4 (available in the online version of this article).

## INTRODUCTION

*Shigella* is a Gram-negative genus comprising four species; *Shigella dysenteriae*, *Shigella boydii*, *Shigella sonnei* and

*Shigella flexneri*, all etiological agents of shigellosis [1]. *Shigella* spp. are highly contagious [2] and primarily transmitted via the faecal-oral route through direct person-person contact or contaminated food or water [3]. Presence of a large virulence plasmid (pINV) encoding a type III secretion system supports *Shigella* invasion of colonic epithelial cells, generation of localised inflammation and necrosis, which combined with expression of genes on chromosomal pathogenicity islands (PAIs) generates characteristic symptoms of abdominal cramps and mucoid diarrhoea [4, 5]. Shigellosis accounts for a significant proportion of the global diarrhoea burden; despite mortality decreases since the 1990s, ~125 m estimated cases occur annually, mainly in children <5 years in Asia and Africa [6, 7].

Received 18 January 2021; Accepted 28 June 2021; Published 24 August 2021

**Author affiliations:** <sup>1</sup>University of Bath, Claverton Down Campus, Bath, UK; <sup>2</sup>Gastrointestinal Reference Services, Public Health England, London, UK; <sup>3</sup>Division of Infection and Immunity, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, EH25 9RG, UK.

**\*Correspondence:** Rebecca K. Locke, rkl30@bath.ac.uk

**Keywords:** antimicrobial resistance; CTX-M; ESBL; MSM; public health; *Shigella*.

**Abbreviations:** AMR, antimicrobial resistance; ARG, antimicrobial resistance gene; BAM, binary alignment map; BRIG, BLAST Ring Image Generator; CARD, Comprehensive Antimicrobial Resistance Database; ESBL, extended spectrum beta-lactamase; IS, insertion sequence; MDR, multidrug resistant; MDRE, Multiple Drug Resistance Element; MGE, mobile genetic element; MLST, Multi-Locus Sequence Typing; MSM, men who have sex with men; ONT, Oxford Nanopore Technologies; PAI, pathogenicity island; PHE, Public Health England; QRDR, quinolone resistance determining region; SNP, single nucleotide polymorphism; SRL, *Shigella* Resistance Locus; TIR, terminal inverted repeat; TSD, target sequence duplication.

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

000644 © 2021 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

Historically, *Shigella* infections in higher-income countries such as the United Kingdom (UK) were mainly associated with travellers returning from endemic areas [8]. However, since 2009 the UK has observed sustained epidemics of domestically-acquired multidrug resistant (MDR) *Shigella* serotypes *S. flexneri* 3a and 2a and *S. sonnei*, attributed to transmission in men who have sex with men (MSM) [9]. Ciprofloxacin, azithromycin, and ceftriaxone are now first and second-line treatment options [10] due to chromosomal acquisition of antimicrobial resistance genes (ARGs) relevant to earlier drugs. This includes the class II integron Tn7/Int2 (trimethoprim and aminoglycosides), class I *Shigella* Resistance Locus (SRL) (chloramphenicol, ampicillin and tetracycline) and additional sulphonamide resistance via the small spA plasmid [11–13]. However, ciprofloxacin use is restricted by step-wise chromosomal mutations in the quinolone resistance determining regions (QRDR) *gyrA* and *parC* [14] and intercontinental dissemination of azithromycin ARGs on a single large IncF plasmid, pKSR100, has directly enhanced previous MSM *Shigella* outbreaks [15].

Extended spectrum  $\beta$ -lactamase (ESBL) genes have implications for failure of ceftriaxone, one of the few remaining options, and were relatively rare in domestically-acquired *Shigella* in the UK [8]. However, in 2015 Public Health England (PHE) described an ESBL-producing *S. sonnei* MSM cluster [16] conferred by a pKSR100-like plasmid (p183660), which had acquired *bla*<sub>CTX-M-27</sub>. This element recently drove a prolonged *S. sonnei* outbreak among MSM in Victoria, Australia [17] but few CTX-M *Shigella* plasmids have been characterised [18]; seemingly only one other has been described in detail, in a Swiss *S. sonnei* isolate [19].

Assembly of these plasmids is important for AMR surveillance but is undermined by the failure of short-read sequencing to span repetitive mobile genetic element (MGE) regions, preventing ARG location determination [20]. Advent of long-read sequencing such as the Oxford Nanopore Technologies (Nanopore) platform, in combination with Illumina-mediated polishing, overcomes this to provide accurate genomes with complete plasmids [21, 22].

In this study we report comparative genomics of four MSM-associated *Shigella* strains (three *S. sonnei*, one *S. flexneri* 3a) in England initially characterised as *bla*<sub>CTX-M-27</sub>-positive. Availability of complete ESBL-producing *Shigella* genomes helps determine structural MGE arrangement, genomic ARG context and identify insertion sequences (IS) involved in their reorganisation. Together this helps elucidate how MGE evolution is tied to pathogenic persistence, emphasising the importance of responsible antimicrobial stewardship.

## EXPERIMENTAL PROCEDURES

### Short-read sequencing (Illumina HiSeq 2500) and data processing

Four strains of *Shigella* spp. (*Shigella sonnei*, *n*=3, *Shigella flexneri* 3a, *n*=1) from MSM were selected as representatives of clusters of all *Shigella* species in the PHE database with

### Impact Statement

Outbreaks of *Shigella*, a gastrointestinal pathogen, have been known to occur in men who have sex with men (MSM) globally since the 1970s. Increasing prevalence of multidrug resistance in *Shigellae* is concerning due to the challenge this poses to current antimicrobial treatment options. This includes the emergence of extended-spectrum  $\beta$ -lactamase (ESBL)-producing strains which are resistant to ceftriaxone due to plasmid-mediated CTX-M determinants. As few CTX-M plasmids have been described, we apply a combination of short- and long-read sequencing with Illumina and MinION technologies to assemble complete plasmids in four *Shigella* isolates associated with MSM transmission in the United Kingdom to investigate the location of antimicrobial resistance (AMR) genes. We find *bla*<sub>CTX-M-27</sub> is inserted into plasmids with high identity to known plasmids which have previously driven shigellosis epidemics worldwide and typically carry azithromycin resistance elements. We compare isolate differences in AMR gene context, the impact of insertion sequences on their plasmids and discuss how resistance profiles relate to MSM clade. Our findings highlight the necessity of increased public health monitoring of AMR in *Shigella* at different hierarchies; the whole genome, mobile genetic element and single gene level to understand resistance dissemination and inform effective treatment options.

the *bla*<sub>CTX-M-27</sub> determinant and characterised in this study (Table 1). If the isolate formed part of a cluster, the earliest isolate was selected with respect to faecal sample collection date (or date of isolation if sample data were unavailable). For Illumina sequencing, Genomic DNA was extracted from cultures using the QIAasympyphony system (Qiagen). The sequencing library was prepared using the Nextera XP kit (Illumina) for sequencing on the HiSeq 2500 instrument (Illumina), run with the fast protocol. FASTQ reads were processed using Trimmomatic v0.27 [23] to remove bases with a PHRED score of <30 from the leading and trailing ends, with reads <50 bp after quality trimming discarded.

### Whole genome long-read sequencing (Nanopore) and data processing

Samples were subsequently sequenced using Oxford Nanopore technologies. Genomic DNA was extracted and purified using the Fire Monkey DNA extraction kit (Revolutgen) with subtle modifications to manufacturer's instructions including removal of vortexing steps. Genomic DNA for each extract was quantified using a Qubit and the HS (high sensitivity) dsDNA assay kit (ThermoFisher Scientific), following manufacturer's instructions. Library preparation was performed using the Rapid barcoding kit SQK-RBK004 (Oxford Nanopore Technologies). The prepared libraries were loaded onto

**Table 1.** Epidemiological information related to the four *Shigella* isolates associated with men who have sex with men (MSM) that were whole genome sequenced with Nanopore and Illumina technologies in this study. *S. sonnei* MSM clade is reported according to [8] and *S. sonnei* lineage is included according to [47]

Isolate ID	598080	607387	888048	893916
Species	<i>S. sonnei</i>	<i>S. sonnei</i>	<i>S. flexneri</i> 3a	<i>S. sonnei</i>
Region	London	London	London	London
Sex	M	M	M	M
Age	22	50	50	36
Receipt date (M/Y)	08/18	09/18	02/20	02/20
<i>S. sonnei</i> MSM clade	2	2	–	5
<i>S. sonnei</i> lineage	3.7.29.1.4.1 (Global III VN2.KH1.Aus)	3.7.29.1.4.1 (Global III VN2.KH1.Aus)	–	3.6.1.1.2 (CipR.MSM5)

a FLO-MIN106 R9.4.1 flow cell (Oxford Nanopore Technologies) and sequenced using the MinION (Oxford Nanopore Technologies) for 36 h and further processed as follows (Fig. S1). Data produced in a raw FAST5 format was basecalled and de-multiplexed with Guppy v3.2.10 using the FAST protocol (Oxford Nanopore Technologies) into FASTQ format and de-multiplexed into each samples' respective barcode.

Read metrics prior to and following filtering were determined with NanoPlot v1.8.1 [24]. Adapter sequences were trimmed from raw Nanopore reads and chimeric reads discarded to prevent cross-barcode contamination using Porechop v0.2.4 (<https://github.com/rrwick/Porechop>). Read filtering was conducted with Filtlong v0.2.0 (<https://github.com/rrwick/Filtlong>) to yield 30× theoretical coverage of the *Shigella* genome (~4.7 Mb) with the longest reads using the following parameters; min\_length=1000, keep\_percent=90, length\_weight=10, target\_bases=141 Mb.

### De novo assembly, polishing and reorientation

All genomes were *de novo* assembled from filtered ONT FASTQ files and two assemblers were compared for each strain. First, long-read assembly was performed with Flye v2.7.1 [25] with an input genome size of 4.7 Mb and use of the --plasmids parameter to rescue any small unassembled replicons. A hybrid assembly combining paired-end Illumina and Nanopore reads was also generated with Unicycler v0.4.8 [26]. The resulting assemblies were compared for contiguity (total contig number and contig N<sub>50</sub>) and plasmid recovery and the Flye-generated assembly was utilised for further investigation in all cases (Table S2). Assemblies were then polished to reduce complex break, insertion and base-level substitution errors; Illumina FASTQ reads were mapped to the draft assembly with the Burrows-Wheeler Aligner (BWA) v0.7.17 MEM algorithm [27] and Samtools v1.7 [28] with the -F 256 parameter to ignore non-primary aligning reads. This was achieved with two rounds of the Pilon v1.23 –fix all function [29] with the following parameters; minimum depth 0.05, minimum quality 30 and minimum mapping quality 30. This was followed by two rounds of Racon v1.4.13 [30] with a match score of 8, average base quality threshold for windows of 30, mismatch score of –6 and gap penalty of

–8. The final assembly quality was determined by QUAST v5.0.2 [31]. Where possible, contigs were reoriented on *dnaA* or *repA* using the Circlator v1.5.5 fixstart function [32]. As one assembly (607387) was poorly resolved, contigs were re-ordered according to the high quality *S. sonnei* 53G reference (GenBank NC\_016822.1) using Mauve v2.4.0 [33].

### Annotation, MLST and plasmid comparison

FASTA files were annotated with Prokka v1.14.6 [34] using a protein reference guide of pKSR100 (Accession LN624486) proteins and those identified at high confidence by ResFinder; in cases of ambiguity proteins were checked manually with the non-redundant BLASTp database. Then mlplasmids v1.1.0 [35] and BLASTn were used to predict chromosomal or plasmid contig origin. Multi-Locus Sequence Typing (MLST) was performed using the Centre for Genomic Epidemiology (CGE) database *E. coli* scheme #1 based on allelic profiles of seven housekeeping genes: *adh*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* and *recA* [36]. IncFII plasmids were compared using BLASTn v2.10.1 with default parameters for percent identity and query cover to pKSR100 (GenBank LN624486.1), a conjugative plasmid from an *S. flexneri* strain SF795513 associated with shigellosis in MSM. As isolates were known to harbour *bla*<sub>CTX-M-27</sub> they were also compared to p183660, found previously in *S. sonnei*, with high identity to pKSR100 and harbouring *bla*<sub>CTX-M-27</sub> (GenBank KX008967.1) [16]. Other plasmids were compared to known plasmids using the non-redundant BLASTn database. ISSaga [37] was used to annotate IS in pKSR100-like plasmids and to detect overall numbers in genomes.

### Antimicrobial resistance and virulence gene identification

The CGE PlasmidFinder-2.0 and pMLST-2.0 Enterobacteriaceae plasmid replicon database [38] was used to detect and type plasmids with default parameters (>95% nucleotide identity threshold and >60% query coverage). Acquired ARGs in *Shigella* samples were detected *in silico* from assembled sequences using the ResFinder-3.2 [39] database including known *E. coli* chromosomal mutations, both with default 90% identity and 80% minimum length thresholds but



all genes found and reported exhibited >98% identity and >99% coverage. The Comprehensive Antimicrobial Resistance Database (CARD) Resistance Gene Identifier tool [40] was also used to identify ARGs and mutations with default parameters (Perfect and Strict hits), as ResFinder does not generally include chromosome-specific genes. To investigate discrepancies in ARG identification, Illumina and Nanopore reads were mapped to *bla*<sub>CTX-M-27</sub> from the CARD database using BWA-MEM v0.7.17 and minimap2 v2.17 [41] respectively with Samtools v1.7 where BAM files were visualised using Tablet v1.19.09.03 [42]. The ISEScan tool v.1.7.2.1 [43] was used to identify terminal inverted repeats associated with IS. Virulence factors were detected within assemblies using VirulenceFinder (<https://cge.cbs.dtu.dk/services/VirulenceFinder>) with default parameters.

### Maximum-likelihood phylogeny

To provide context for the four *Shigella* isolates, maximum-likelihood phylogenies were created for the *S. flexneri* 3a isolate and 49 other *S. flexneri* genomes (50 total) (Table S3), and the *S. sonnei* isolates with 198 other *S. sonnei* strains (201 total) from PHE's *Shigella* genome collection isolated in England. Representative strains from relevant unique 250 single-linkage hierarchical clusters were randomly selected from different time frames and SnapperDB v0.2.6 (get\_the\_snps) [44] was used to generate core variant multiple sequence alignments of 4415 bp and 11096 bp for *S. sonnei* and *S. flexneri* respectively. SnapperDB stores reference-based single nucleotide polymorphism (SNP) variant calls relative to the reference genome, with each isolate given a SNP Address based on single-linkage hierarchical clustering which is a proxy for the genetic distances of isolates in the population. In detail, Illumina reads were aligned to a reference genome, consisting of the *S. sonnei* strain Ss46 (GenBank accession number NC\_007384.1) and *S. flexneri* 2a strain 2457T (GenBank accession number AE014073.1) using BWA-MEM v0.7.12 as previously described [45]. Sequence alignment maps were sorted and indexed to produce a binary alignment map (BAM) file using Samtools v1.0.18. SNPs were identified using GATK v2.6.5 and only high-quality SNP (mapping quality [MQ] >30, minimum depth >10, variant ratio >0.9) positions were extracted. These polymorphic position alignments were used to infer maximum-likelihood phylogenetic trees using IQ-Tree v2.0.6 [46] with 1000 ultra-fast bootstrap approximations. The general time reversible model plus ascertainment bias correction was used to correct branch length overestimation due to absence of constant sites.

### Lineage typing

Discriminatory, lineage specific SNPs were defined based on the phylogenetic lineage assignment in Baker *et al.* [8] and extracted directly from SnapperDB v0.2.5. To ensure greater utility to public health researchers, lineage typing of *S. sonnei* isolates (including publicly available AUSMDU00008333 (GenBank accession ERA1715196) and 183660 (GenBank accession SRX1766927)) was also performed according to a newly described standard scheme [47] with Mykrobe

v0.9.0 (predict) software [48] using the --ont parameter with trimmed FASTQ reads as input.

### Metadata and ARG presence for isolates used for phylogenetic context

Not all isolates included for context were MSM-linked, however shigellosis cases that were (i) male, (ii) adults and (iii) had not reported recent foreign travel were inferred to be likely associated with sexual transmission among MSM. Presence of genes conferring resistance to azithromycin (*erm(B)*, *mph(A)*) and mutations in the quinolone resistance-determining regions (QRDR) of *gyrA* and *parC* for all isolates included in the two phylogenetic trees were extracted from a previously published study [45]. This excludes *S. sonnei* isolate 893916 where ResFinder results were used to determine presence of these genes and mutations. Detection of *bla*<sub>CTX-M-27</sub> within all genomes was determined using a mapping-based approach known as GeneFinder v2.2 ([https://github.com/phe-bioinformatics/gene\\_finder](https://github.com/phe-bioinformatics/gene_finder)) with default parameters. GeneFinder utilises Bowtie2 v2.1 [49] and Samtools v1.0.18 [28] to align Illumina query sample reads to a reference.1.1.

### Visualisation tools

Graphical comparison of chromosomal islands was performed using EasyFig v2.2 [50] and SRL genomic features of *S. flexneri* 3a were visualised with GView 1.7 [51]. BLAST Ring Image Generator (BRIG) [52] with BLAST v2.10.1 was utilised to visualise similarity of plasmids to pKSR100 and p183660, the latter of which was used as a reference. BRIG was also used to compare IncB/O/K/Z plasmids to a plasmid identified with high similarity by BLASTn, pAUSMDU00008333\_3 (GenBank LR213460.1); isolate 607387 was used as a reference due to it being the largest plasmid and virulence plasmids (pINV) were compared to one another. The R package ggtree [53] was used to midpoint root maximum-likelihood trees, annotate tip nodes and visualise trees with isolate metadata and resistance gene presence/absence as an adjacent heatmap.

## RESULTS

### Genomic features and assemblies

Four MSM-associated *Shigella* isolates (three *S. sonnei*; two 2018, one 2020 and one *S. flexneri* 3a, 2020) were selected for long-read sequencing due to *bla*<sub>CTX-M-27</sub> presence in initial Illumina reads (Table 1). In all cases Flye outperformed Unicycler in creating more contiguous assemblies (Tables 2 and S2); detailed contig-level information for each isolate is available in Table S4.

### Virulence determinants and known virulence plasmid, pINV

Detection of virulence factors revealed small differences between strains and no genes encoding Shigatoxin (Table 3). All isolates harboured the large ~220 kbp virulence plasmid pINV (Fig. S5); pivotal for *Shigella* pathogenic evolution through acquisition of

**Table 2.** Assembly statistics, genomic features and predicted number of insertion sequence (IS) elements of the three *S. sonnei* isolates and one *S. flexneri* 3a isolate. Genomes were constructed with Flye, where Illumina short reads were utilised to polish a draft long-read assembly to yield a more complete and accurate final assembly

Isolate ID	598080	607387	888048	893916
Species	<i>S. sonnei</i>	<i>S. sonnei</i>	<i>S. flexneri</i> 3a	<i>S. sonnei</i>
MLST	152	152	245	152
Total contigs	6 (2 Chromosomal contigs, 3 plasmids, 1 likely plasmid contig)	37 (24 likely Chromosomal contigs, 4 plasmids, 9 likely plasmid contigs)	5 (1 Chromosomal contig, 2 plasmids, 2 likely plasmid contigs)	4 (1 Chromosomal contig, 3 plasmids)
Total size (bp)	5282682	5315865	4839032	5196904
Largest contig (bp)	4879016	2500534	4519004	4813904
N <sub>50</sub>	4879016	897518	4519004	4813904
L <sub>50</sub>	1	2	1	1
GC content (%)	50.86	50.82	50.68	50.84
Total CDS	5533	5554	5174	5462
rRNA	22	22	23	22
tRNA	102	101	98	96
tmRNA	1	1	1	1
Acquired resistance genes (ResFinder)	11	9	7	11*
Resistance genes (CARD)†	59	58	47	60
Virulence factors (VFDB)	14	13	12	13
Predicted IS number	577	520	504	588
Estimated different IS	45–48	38–43	43–47	43–44
SRR number (Illumina)	SRR7842065	SRR7892120	SRR11096691	SRR11206407
SNP address	1.3.197.460. 1360.1582. 2770	1.3.197.460. 1360.1582. 2813	35.43.43. 43.1189. 2224.2978	1.1.1.1.377. 394.3945

\*Two mutations in *gyrA* are counted as two genes.

†Perfect and Strict hits only, total includes those also found by ResFinder.

genes such as the conserved 30kbp *mxi-spa* locus encoding a type III secretion system, facilitating host invasion and spread [54]. pINV carried *ipaD*, *capU* and *virF*, with the *S. flexneri* isolate containing an additional *sepA* gene, thought to intensify intestinal fluid accumulation [55]. *S. sonnei* isolates displayed similar virulence profiles to *S. flexneri* but lacked *sat*, instead possessing *lpfA*, *sigA* and *senB*, hallmarks of *S. sonnei* virulence [56].

### Shared chromosomal AMR determinants among *Shigella* species

All isolates were genotypically and phenotypically MDR (resistant to ≥3 antimicrobial classes) and harboured known *Shigella* MGEs (Fig. 1) according to [8] (Table S5). Sequence analysis showed the presence of chromosomal class 2 IntI2/Tn7 integrons in all *S. sonnei* isolates. For 598080 and 607387 this comprised the canonical ARG cassette organisation

*dfrA1-sat2-aadA1* (Fig. 2a), conferring trimethoprim, streptomycin and streptomycin resistance respectively, however 893916 lacked *aadA1*. This was unexpected as IntI2 is unable to alter the array due to an internal stop codon, usually leading to constant arrangements. All *S. sonnei* isolates harboured a 3' *tns* transposition gene segment (*tnsA-E*) inserted adjacent to glutamine-fructose-6-phosphate aminotransferase, *glmS*, suggesting a single acquisition event. No such cassette was present in the *S. flexneri* isolate, which instead harboured *aadA1* on a class one integron, the SRL (Fig. 2b) alongside *bla*<sub>OXA-1</sub>, *catA1* and *tet(B)* genes. These confer resistance to streptomycin, β-lactams, chloramphenicol and tetracyclines alongside a ferric uptake transport system (*fecI-R* with downstream structural *fecABCDE* genes), all integrated into tRNA-Ser.

Ciprofloxacin resistance is conferred by chromosomal Quinolone Resistance Determining Region (QRDR) mutations

**Table 3.** Virulence profiles of four *Shigella* isolates sequenced in this study, as determined by VirulenceFinder from whole genome assemblies. x indicates virulence factor present, – indicates virulence factor absent

Virulence factor	598080 ( <i>S. sonnei</i> )	607387 ( <i>S. sonnei</i> )	888048 ( <i>S. flexneri</i> 3a)	893916 ( <i>S. sonnei</i> )
Invasion protein ( <i>ipaD</i> )	x	x	x	x
Hexosyltransferase homlog ( <i>capU</i> )	x	x	x	x
VirF transcriptional activator ( <i>virF</i> )	x	x	x	x
<i>Shigella</i> extracellular protein A ( <i>sepA</i> )	–	–	x	–
Glutamate decarboxylase ( <i>gad</i> )	x	x	x	x
Invasion plasmid antigen ( <i>ipaH</i> )	x	x	x	x
Aerobactin synthetase ( <i>iucC</i> )	x	x	x	x
Ferric aerobactin receptor ( <i>iutA</i> )	x	x	x	x
Secreted autotransporter toxin ( <i>sat</i> )	–	–	x	–
Iron transport protein ( <i>sitA</i> )	x	x	x	x
Tellurium ion resistance protein ( <i>terC</i> )	x	x	x	x
Outer membrane complement resistance protein ( <i>traT</i> )	x	x	x	x
Long polar fimbriae ( <i>lpfA</i> )	x	x	–	x
<i>Shigella</i> IgA-like protease homolog ( <i>sigA</i> )	x	x	–	x
Enterotoxin ShET-2 ( <i>senB</i> )	x	x	–	x
Endonuclease colicin E2 ( <i>celb</i> )	x	x	–	–

[8]. The two 2018 *S. sonnei* strains harboured single *gyrA* mutations (S83L) conferring reduced susceptibility to fluoroquinolones and are concerning as this can act as a prerequisite for successive stepwise mutations [57], implying possible evolution into fully resistant clones. Contrastingly, the 2020 *S. sonnei* strain had triple mutations (*gyrA* S83L; D87G, *parC* S80I) conferring high-level resistance. No QRDR mutations were found in *S. flexneri* 3a and, overall, no plasmid-encoded quinolone ARGs were detected.

All isolates harboured chromosomal class C  $\beta$ -lactamases *ampC* (between *frdD* and *blc* genes) and *ampH* (between *sbmA* and *iprA*) (Fig. 1b) promoting cephalosporin resistance [58]. These results are all concordant with phenotypic data, excluding the finding that 893916 was phenotypically resistant to tetracyclines, despite no *tet* genes being found by ResFinder or CARD (Figs 1b and S2).

### Phylogenetic analysis and clustering within MSM clades

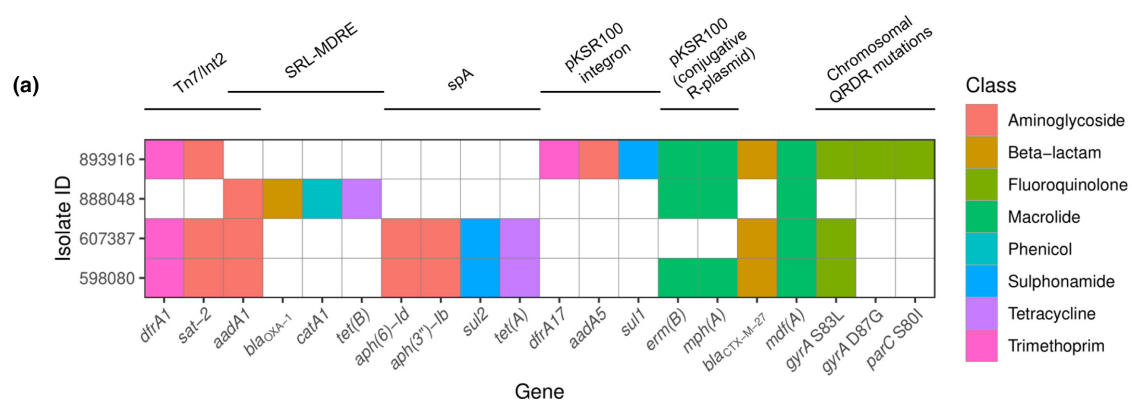
To determine evolutionary relationships of strains within *Shigella*, we undertook separate analyses of 50 *S. flexneri* isolates and 201 *S. sonnei* strains based on SNP differences. The *S. flexneri* 3a isolate clustered within an *S. flexneri* 3a clade with high genetic homogeneity that emerged in 2015 and persists in the MSM population to date (Fig. 3), suggesting a sustained outbreak with possibly one transmission chain. Previous phylogenetic analysis has shown decreases in *S. flexneri* 3a cases and effective population size over time in English MSM, reflected in decreased sample diversity [45] as displayed here. Isolate 888048 was the only *S. flexneri* isolate

with presence of *bla*<sub>CTX-M-27</sub>, within a clade where almost all isolates harbour single *parC* mutation S57R and the majority contain both *erm(B)* and *mph(A)*, though some isolates exhibit neither macrolide resistance determinant or *mph(A)* alone.

*S. sonnei* is divided into four lineages, one of which (lineage III) disseminated globally after MDR acquisition [11], and comprises five MSM-specific clades [8]. The two 2018 *S. sonnei* isolates (green and pink triangles) clustered within MSM clade 2 alongside mostly 2018 isolates (Fig. 4). This clade is known to have acquired pKSR100 (with *erm(B)* and *mph(A)*) on multiple occasions, have single QRDR mutations and links to Australian MSM but has a declining effective population with stable case rates [45]. Lineage typing according to a new scheme denotes 598080 and 607387 to be part of the VN2. KH1.Aus lineage (genotype 3.7.29.1.4.1); Australian isolates that emerged from the Khanh Hoa region of Vietnam [47].

Contrastingly, the 2020 isolate clustered within MSM clade 5 isolates collected from 2017 to 2020. Members are known to harbour pKSR100 and concerning triple QRDR mutations (some exhibited an additional *parC* E84G mutation) and have an increasing effective population size since 2017 [45]. When utilising the standard typing scheme, 893916 was part of the CipR. MSM5 sublineage (genotype 3.6.1.1.2), which emerged from South Asia in the early 2000s, is linked to MSM communities and is responsible for the majority of fluoroquinolone resistant *S. sonnei* infections in Australia, England and the USA [47].

We found *bla*<sub>CTX-M-27</sub> is not restricted to one *S. sonnei* MSM clade and may have been acquired on multiple occasions;



(b)

Isolate ID	Ampicillin	Ceftriaxone	Fluoroquinolones	Macrolides	Trimethoprim	Tetracyclines	Sulphonamides	Chloramphenicol
893916	R	R	R	R	R	R	R	S
888048	R	S	S	R	S	R	S	R
607387	R	R	I	S	R	R	R	S
598080	R	R	I	R	R	R	R	S

**Fig. 1.** (a) Binary heatmap of antimicrobial resistance genes (ARGs) and chromosomal mutations present in four *Shigella* strains identified by ResFinder-3.2; *S. sonnei* (598080, 607387 and 893916) and *S. flexneri* 3a (888048). Where a gene is present (at a threshold of >98% identity and >99% query coverage), the tile is coloured by the antimicrobial class it confers resistance to. The gene *sat-2* is not included in the ResFinder database but is included in this figure and chromosomal mutations in quinolone resistance determining region (QRDR) genes *gyrA* and *parC* are shown on the right. Known mobile genetic elements in *Shigella* are denoted above, as previously defined by Baker et al. [8]. (b) Phenotypic resistance profiles of the four isolates to antimicrobial classes. R=Resistant, S=Susceptible, I=Intermediate.

893916 is the only MSM clade 5 isolate with this element whereas it is present in all MSM clade 2 and only one MSM clade 4 isolate (Fig. 4). *bla*<sub>CTX-M-27</sub> is only found in isolates with both *erm(B)* and *mph(A)* (likely co-located on the same plasmid) or neither determinant. Clustering of non-MSM cases within MSM clades could indicate community acquisition from MSM (e.g. shared households), or onward transmission to MSM from travel-related cases.

### ***S. sonnei* harbour AMR determinants on IncB/O/K/Z plasmids**

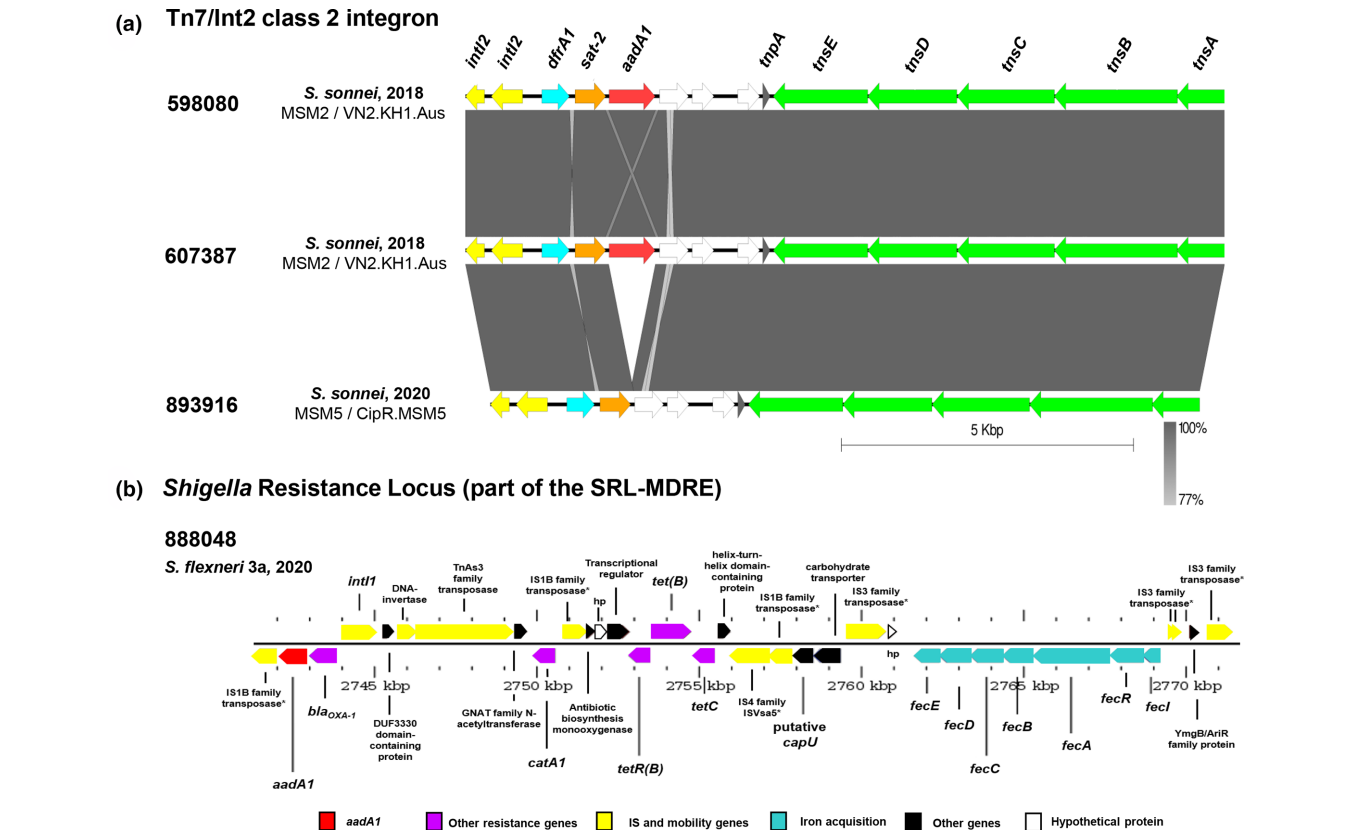
PlasmidFinder revealed a large IncB/O/K/Z plasmid (86–103 kbp) in all *S. sonnei* isolates with high nucleotide similarity (96.3–99.5% identity/79–91% coverage) to a plasmid circulating in Australian MSM from a 2 year population-level study (2016–18), AUSMDU00008333 plasmid 3 [59] (Figs 5 and S4). Compared to this 2016 reference, 598080 and 607307 share an AMR region harbouring resistance to aminoglycosides (*aph(6)-Id* and *aph(3')-Ib*) and sulphonamides (*sul2*), however the two 2018 *S. sonnei* isolates had acquired additional tetracycline *tetR/tet(A)* genes. Notably, the Australian MSM plasmid harboured *bla*<sub>TEM-1c</sub>, but this element was not present in this study and the plasmid within 893916 did not carry AMR determinants, suggesting other genes present on this plasmid are sufficient to outweigh any reproductive fitness cost.

### **Analysis of IncFII plasmids and genomic context of the *bla*<sub>CTX-M-27</sub> gene**

pMLST revealed all four strains harboured a large IncFII ([F2:A:B-]) plasmid (67–83kbp). They demonstrated high similarity to known plasmid pKSR100 (98–99.5% identity/80–90% cover), which has facilitated azithromycin resistance dissemination between *Shigella* MSM lineages [15]. Since isolates were known to harbour *bla*<sub>CTX-M-27</sub> they were compared to p183660, a pKSR100-like MSM plasmid which has acquired this element [16], revealing 99.7–99.9% identity/97–98% cover. These were also compared to each other to understand the organisation and loss of respective ARGs (Fig. 4).

All IncFII plasmids contained a conserved ~35 kbp conjugation-associated *tra* region allowing broader host dissemination, *pemKI* toxin-antitoxin system and maintenance genes. Interestingly, ARGs were located on mosaic regions displaying unique structures for each plasmid (Fig. 6). No single ARG was found ubiquitously and no strains contained non-ESBL  $\beta$ -lactamase gene *bla*<sub>TEM-1</sub>, present in both references pKSR100 and p183660. Three-quarters of isolates harboured *erm(B)* and *mph(A)*; the two 2018 *S. sonnei* isolates (598080 and 607387) had similar resistance profiles overall, differing only in loss of both genes by 607387.



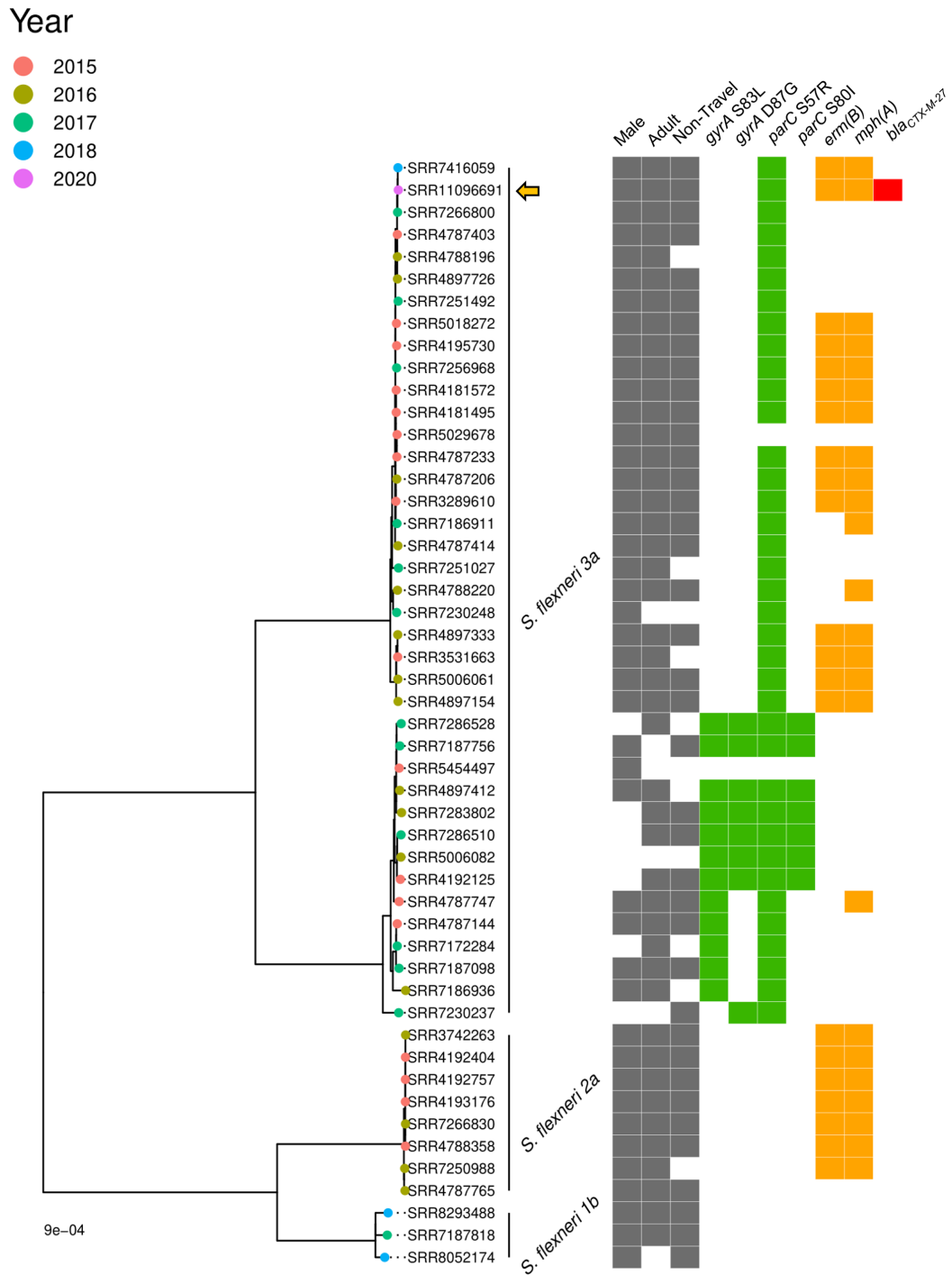


**Fig. 2.** Chromosomal resistance islands identified in four MSM-associated *Shigella* isolates through sequencing with Oxford Nanopore Technologies. (a) BLAST comparisons of Tn7/Int2 integron antibiotic resistance gene cassettes found in all strains of *S. sonnei*. The three have similar arrangements, with 893916 lacking the 3' *aadA1* aminoglycoside resistance gene. *IntI2*, an integrase, is shown in yellow and is non-functional. Genes *dfrA1*: dihydrofolate reductase (blue); *sat-2*: streptothricin acetyltransferase (orange); *aadA1*: aminoglycoside adenylyltransferase (red); encoding resistance to trimethoprim, streptothricin and streptomycin respectively. Hypothetical proteins are shown in white and Tn7 transposition genes shown in green. The direction of the arrow shows the direction of gene coding and grey gradient legend shows BLAST identity between sequences. Figure produced with Easyfig v2.2.5 [50]. (b) Genomic organisation of the chromosomal *Shigella* Resistance Locus, part of the Multiple Drug Resistance Element (SRL-MDRE) present only in the *S. flexneri* 3a strain 888048. This harbours *aadA1*, *bla*<sub>OXA-17</sub>, *catA1* and *tet* genes encoding resistance to aminoglycosides, ampicillin, chloramphenicol and tetracycline respectively. *aadA1* is shown in red due to commonality to Tn7/Int2, other resistance genes are shown in purple. The iron acquisition system is shown in turquoise, mobility related genes in yellow, other genes in black and hypothetical proteins in white. Figure produced with Gview [51].

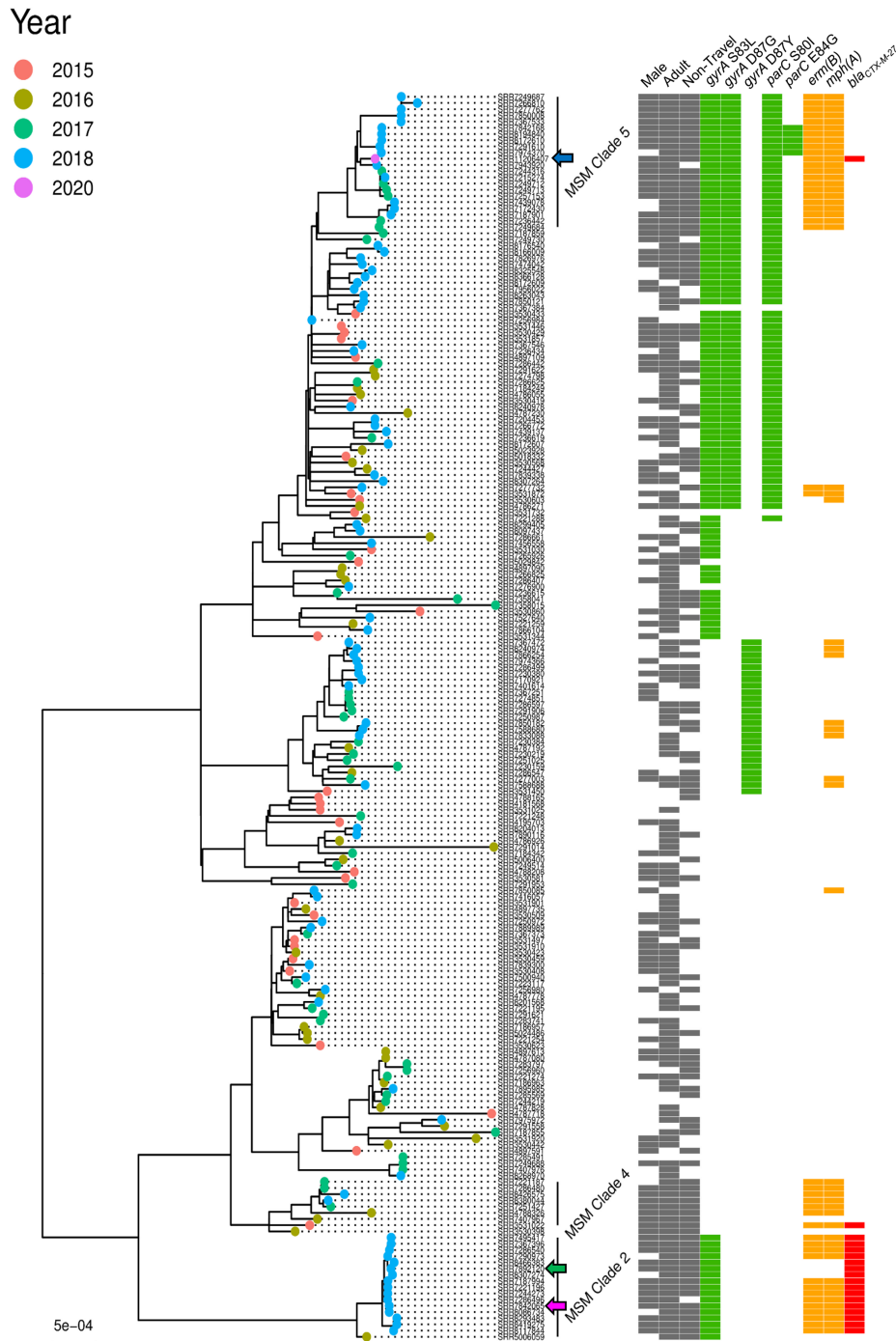
Where present, *erm(B)* was adjacent to its leading peptide (*erm(B)(L)*) and *mph(A)* to its regulatory genes, associated with IS26 as part of an IS26-*mph(A)*-*mrx*-*mphR(A)*-IS6100 unit, though IS6100 was not present in *S. flexneri*. Downstream of *erm(B)*, these genes were always associated with a putative rRNA adenine methyltransferase, a hypothetical protein, a 60 kDa chaperonin (*groL*) and IS91 (Fig. 6).

The *bla*<sub>CTX-M-27</sub> gene (876 bp) was found on all *S. sonnei* plasmids but despite initial Illumina-based *bla*<sub>CTX-M-27</sub> detection in *S. flexneri* 3a (888048), it was not observed in the assembly (Fig. 1). Mapping of trimmed Nanopore reads to a *bla*<sub>CTX-M-27</sub> reference (GenBank AAO61597.1) showed it was not present in these reads (Fig. S3). As this assembly harboured a similar IncFII plasmid with *mph(A)* and *erm(B)*, this suggests *bla*<sub>CTX-M-27</sub> loss during storage or subculture in the 2 month duration between Illumina and Nanopore sequencing (Table S1),

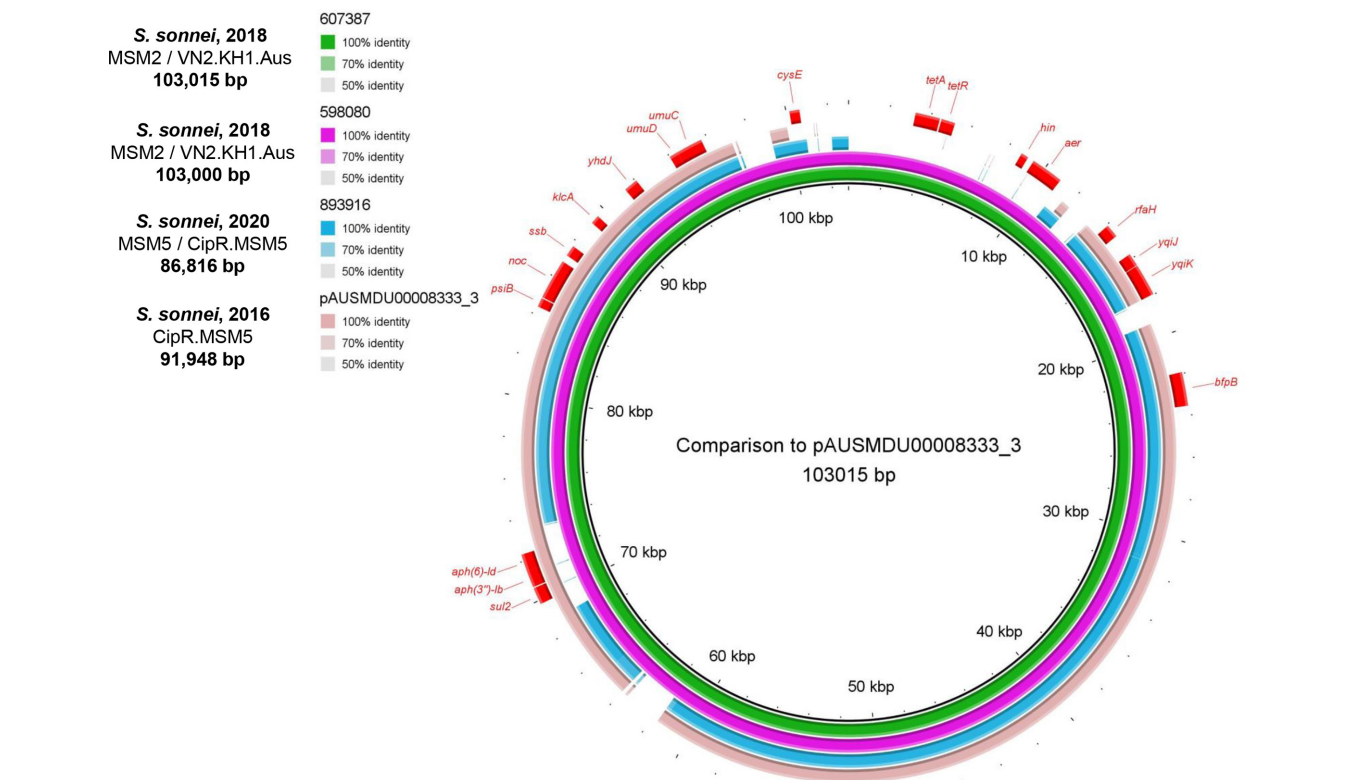
rather than whole plasmid loss or assembly errors (though it is possible *bla*<sub>CTX-M-27</sub> was present on another plasmid that was lost during culture or through mis-assembly). Where present, *bla*<sub>CTX-M-27</sub> was flanked upstream by IS26 and downstream by IS903B. Isolate 893916 was the only one to harbour the previously described pKSR100 integron present in p183660 with sulphonamide, trimethoprim and aminoglycoside resistance genes [*sul1*/*dfrA17*/*aadA5*] alongside *emrE* (*qacEdelta1*), quaternary ammonium compound-resistance protein. This region included heavy metal resistance protein (*srpC*) and vitamin B12 transporter (*btuB*) following *bla*<sub>CTX-M-27</sub>. As IS91 is found between the pKSR100 integron and *mph(A)-erm(B)* unit in 893916 but is lost in 888048 which lacks this integron while harbouring *mph(A)-erm(B)*, despite being present in 598080 with the same context, it is not clear which region IS91 is associated with (Fig. 6).



**Fig. 3.** Maximum-likelihood phylogeny of *S. flexneri* 3a isolate 888048 (SRR11096691, indicated by a filled orange arrow) sequenced with Nanopore technologies in this study, compared to other *S. flexneri* isolates in Public Health England's collections. To place the isolate in context, 49 other *S. flexneri* isolates that are representatives from relevant SNP (single nucleotide polymorphism) clusters were included in the comparison. MSM clades associated with *S. flexneri* 1b, *S. flexneri* 2a and *S. flexneri* 3a are labelled. Isolates are labelled by SRR number and tip nodes are coloured by year of receipt. Adjacent grey blocks in each of the first three columns indicate the sample was isolated from a male, an adult, and that the illness was non-travel related (left-right). Grey blocks in these three lanes for an isolate implies the infection may be associated with MSM transmission. Presence of antimicrobial resistance genes concerning fluoroquinolones (mutations in *gyrA* and *parC*, green) and macrolides (*erm(B)* and *mph(A)*, orange) are taken from Bardsley *et al.* [45] and presence of *bla*<sub>CTX-M-27</sub> was determined using GeneFinder. Gene presence is indicated by a coloured tile in the relevant column. ML trees are midpoint rooted and scale bar indicates SNPs.



**Fig. 4.** Maximum-likelihood phylogeny of three *S. sonnei* isolates known to be associated with men who have sex with men that were sequenced using Nanopore technologies in this study, compared to other isolates in Public Health England's collections. To place the isolates in context, 198 other *S. flexneri* isolates that are representatives from relevant SNP (single nucleotide polymorphism) clusters were included in the comparison. *S. sonnei* isolates investigated in detail in this study are as follows; 598080 (SRR7842065, pink filled arrow), 607387 (SRR7892120, green filled arrow), 893916 (SRR11206407, blue filled arrow). Isolates are labelled by SRR number and tip nodes are coloured by year of receipt. *S. sonnei* MSM clades are labelled as per those described in Baker et al. [8]. Adjacent grey blocks in each of the first three columns (left-right) indicate the sample was isolated from a male, an adult, and that the illness was non-travel related (left-right). Grey blocks in these three lanes for an isolate implies the infection may be associated with MSM transmission. Presence of antimicrobial resistance genes concerning fluoroquinolones (mutations in *gyrA* and *parC*, green) and macrolides (*erm(B)* and *mph(A)*, orange) are taken from Bardsley et al. [45] and presence of *bla<sub>CTX-M-27</sub>* was determined using GeneFinder. Gene presence is indicated by a coloured tile in the relevant column. ML trees are midpoint rooted and scale bar indicates SNPs.



**Fig. 5.** BLAST comparisons of IncB/O/K/Z plasmids for three MSM *S. sonnei* isolates in England, obtained with Nanopore sequencing. These are compared to a plasmid known to be associated with MSM in Australia, pAUSMDU00008333 plasmid 3 (LR213460.1). Three *S. sonnei* plasmids are from this study (isolates 598080; pink ring, 607387; green ring and 893916; blue ring) and the outer peach coloured ring is pAUSMDU00008333 plasmid 3. Genes present are shown in red on the outer ring (excluding hypothetical proteins). *aph(3')-Ib* (*strA*): Aminoglycoside 3'-phosphotransferase; *aph(6)-Ia* (*strB*): Aminoglycoside O-phosphotransferase; *sul2*: Dihydropteroate synthase; *tet(A)*: Tetracycline efflux protein; *psiB*: Plasmid SOS inhibition protein B; *noc*: Nucleoid occlusion protein; *ssb*: Single-stranded DNA-binding protein; Anti-Restriction protein; *yhdJ*: DNA adenine methyltransferase; *umuD* and *umuC* - UV mutagenesis and repair proteins; *cysE*: Serine acetyltransferase; *hin*: DNA-invertase; *aer*: Aerotaxis receptor; *rfaH*: Transcription antitermination protein; *yqiJ* and *yqiK*: Inner membrane proteins; *bfpB*: Outer membrane lipoprotein. Figure produced with BLAST Ring Image Generator (BRIG) [52] and gradient of BLAST identity is shown in the legend.

## Reorganisation of IncFII plasmids is likely driven by IS26

IS elements drive genomic evolution and typically comprise a transposase, catalysing enzymatic mobility, with flanking short terminal inverted repeats (TIRs) (transposase recognition sites) [60]. ISSaga revealed 577, 520 and 588 total predicted IS in *S. sonnei* genomes (598080, 607387 and 893916 respectively) and 504 in *S. flexneri* 3a (Table 2); lower copy numbers in 607387 are likely due to breaks between contigs. The most frequently identified IS family was IS1, accounting for 28.4, 26.3 and 30.8% in *S. sonnei* and 28.9% in *S. flexneri*. Chromosomal IS family proportions are similar between strains (Fig. 7a) but only *S. flexneri* harbours IS66 and remarkably, no copies of the IS6 family are present chromosomally in any isolates, yet dominate on pKSR100-like plasmids; mainly IS26 (Fig. 7b).

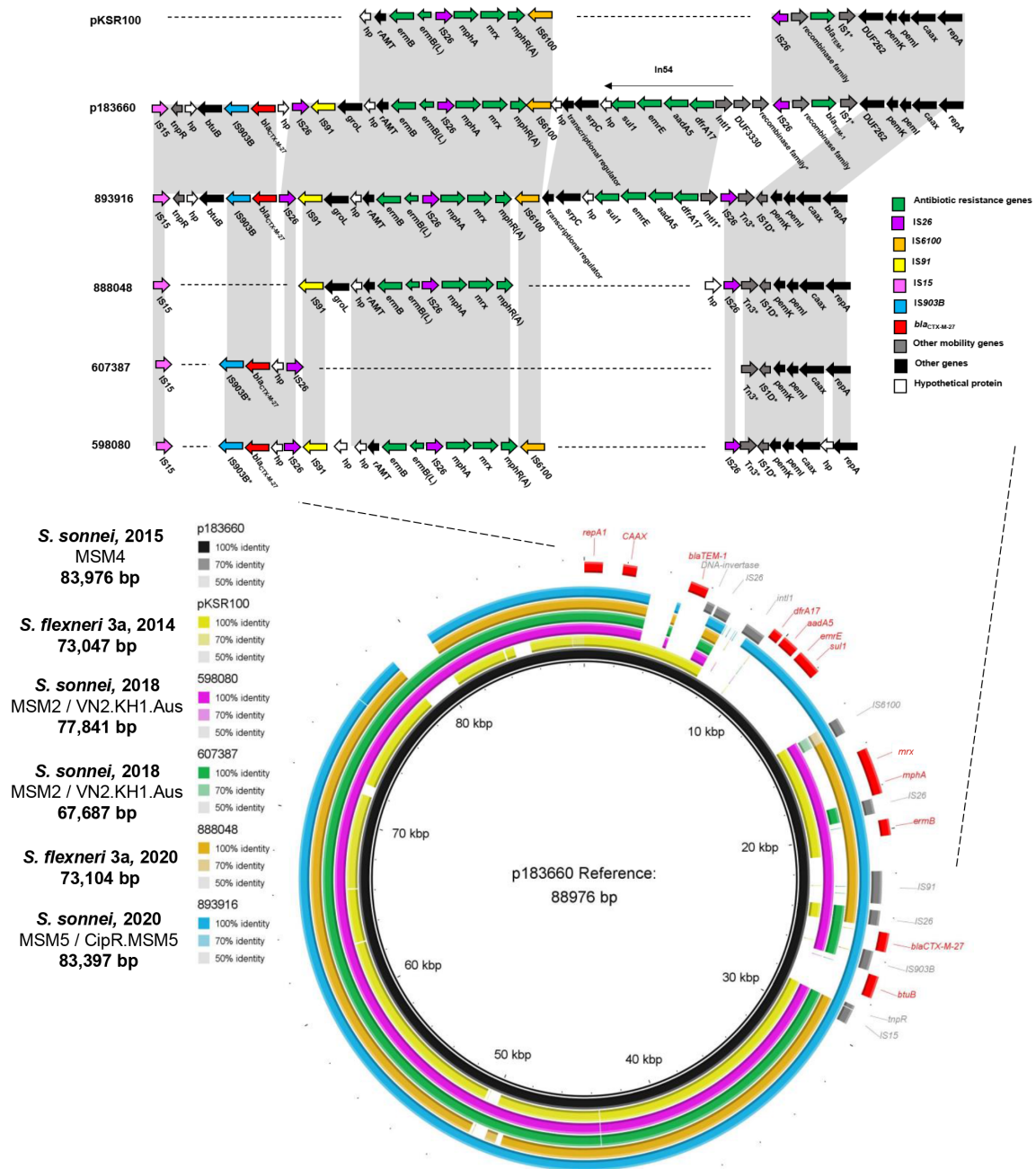
IS26 and IS15 were always flanked by known 14 bp TIR (5'-GGCACTGTTGCAA-3') [61], suggesting possible

recent acquisition without amelioration. When IS26 integrates, in some circumstances this results in 8 bp target sequence duplication (TSD) and flanking TIRs. No TSDs were manually identified so further study is needed to understand translocation mechanisms and acquisition dynamics. IS concentration on these plasmids and flanking of lost regions by IS suggests possible IS-mediated deletion and microevolution driven by IS insertion into the plasmid backbone.

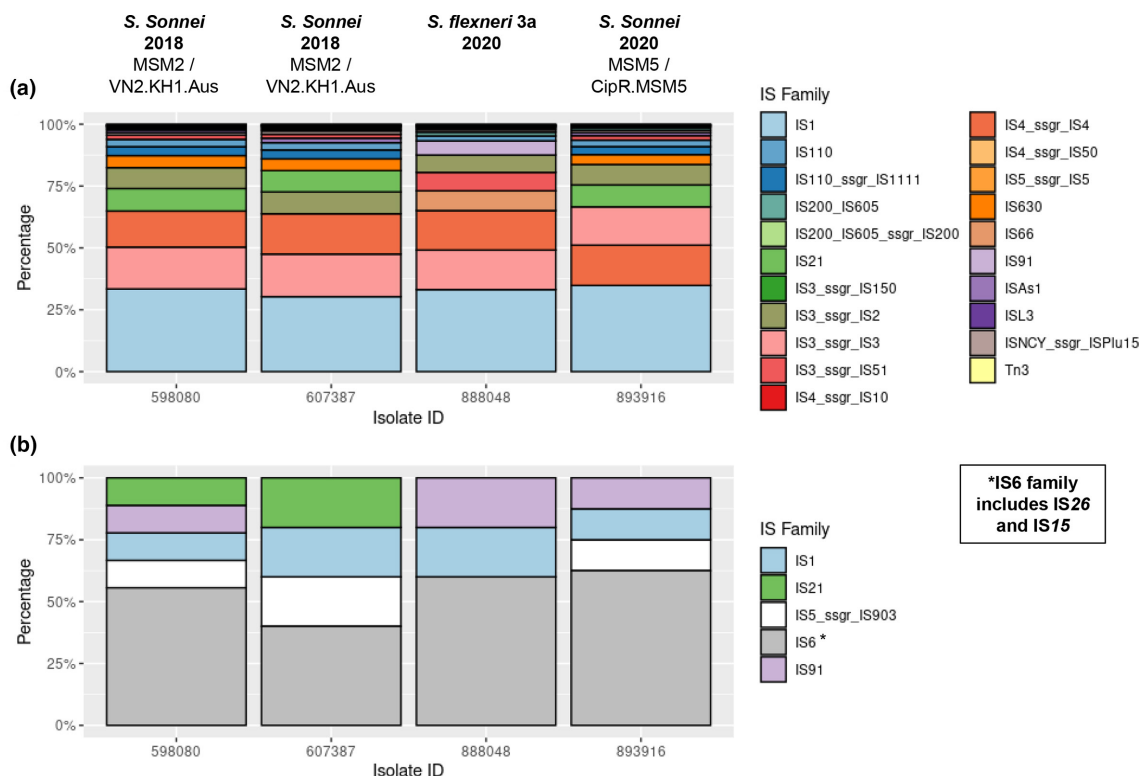
## DISCUSSION

Overall our data show long-read sequencing supports elucidation of genomic AMR context among MSM-associated *Shigella* strains. This includes ESBL (*bla*<sub>CTX-M-27</sub>), borne on a mosaic ARG region within IncFII plasmids which are likely to have been reorganised by IS26. This, together with their phylogenetic position within MSM clades, can aid understanding of how AMR is tied to persistence in the population.





**Fig. 6.** BLAST comparisons of IncFII pKSR100-like plasmids in four *Shigella* isolates from men who have sex with men (MSM) in England obtained with Nanopore sequencing. Three plasmids are from *S. sonnei* (isolates 598080: pink ring; 607387: green ring and 893916: blue ring) and one from *S. flexneri* 3a (888048: orange ring). The second inner ring (yellow) is pKSR100, a known MSM-related plasmid from an *S. flexneri* 3a infection and the innermost (black) ring is p183660, obtained from a man in England infected with *S. sonnei*. This has high identity to pKSR100 but had acquired the pKSR100 integron [*dfrA17*, *aadA5*, *emrE*, *sul1*] and a novel mobile element harbouring *bla*<sub>CTX-M-27</sub>. Upper: DUF- denotes domain-containing proteins. \* Denotes partial features. Features are not drawn to scale. Lower: Sequences associated with mobility in this region are coloured in grey on an outer ring and antimicrobial resistance and other genes are shown on the outer ring in red. *repA1*: IncFII *repA* protein; CAAX: CPBP family intramembrane metalloprotease; *bla*<sub>TEM-1</sub>:  $\beta$ -lactamase; *int11*: Class one integrase; *dfrA17*: Type I dihydrofolate reductase; *aadA5*: Streptomycin adenylyltransferase; *emrE*: Multidrug transporter; *sul1*: Dihydropteroate synthase; *mrx*: Multidrug efflux pump; *mphA*: Aminoglycoside phosphotransferase; *ermB*: Dimethyladenosine transferase; *bla*<sub>CTX-M-27</sub>: ESBL class A  $\beta$ -lactamase, CTX-M-27 type; *btuB*: Vitamin B12 transporter. Figure produced with BLAST Ring Image Generator (BR



**Fig. 7.** The predicted distribution of each IS family identified by ISSaga for each of the four *Shigella* isolates; *S. sonnei* (598080, 607387 and 893916) and *S. flexneri* 3a (888048), by total copy number of each family. *S. sonnei* lineage is shown below isolation year where applicable. (Note: in some cases there are >1 different IS per family). (a) Distribution of each family within contigs identified as chromosomal sequences for each isolate. (b) Distribution of each IS family within the IncFII pKSR100-like plasmid present within each isolate.

Combinations of mutations and presence of ARGs within isolates in this study promote resistance to all current first-line antimicrobials (macrolides, fluoroquinolones and third-generation cephalosporins) [62]. *S. sonnei* isolates harboured Tn7/Int2; this facilitated worldwide dissemination of global lineage III by allowing adaptation to multiple antibiotics in the environment [11]. Only the *S. flexneri* isolate harboured the SRL-MDRE with *bla*<sub>OXA-1</sub> conferring ampicillin resistance, congruent with an Iranian study which suggested *S. flexneri* *bla*<sub>OXA-1</sub> host preference [63]. The *S. flexneri* isolate harboured secretion autotransporter toxin *sat* and *sepA* which is associated with intense abdominal pain [64]. However, *S. sonnei* isolates exhibited higher numbers of virulence genes, including *senB* and *sigA* associated with bloody diarrhoea and fever respectively [65]. Presence of isolates with high levels of both AMR and virulence factors could be a consequence of clinically severe cases requiring antibiotics and exerting selective pressure and are a public health concern due to the challenge this poses to outbreak management.

Reduced susceptibility to azithromycin (RSA) in *Shigella* was not present in England in 2002 [66], but has arisen possibly due to pressure from treatment of other STIs in MSM [67] and triple QRDR mutations promoted emergence of fluoroquinolone-resistant *S. sonnei* in South Asia which

later transmitted internationally [14]. ESBL prevalence is increasing, especially in Asia, North America and Europe [68]; WGS of 335 *S. sonnei* isolates associated with domestic acquisition in England and Wales (2015–16) denotes CTX-M genes in 12% of isolates and *bla*<sub>CTX-M-27</sub> in 2.7% of those, though CTX-M-15 was most common [69]. Use of ceftriaxone to treat shigellosis should be carefully considered, especially in areas with RSA, due to possible co-selection of ESBL with macrolide resistance genes on the same plasmid [15, 59]. There are risks that ESBL-producing strains may spread among HIV-positive MSM with increased biologically susceptibility; use of networking applications is a transmission risk factor [70] so could be utilised for awareness campaigns and contact tracing, as shown effectively during a Berlin MSM hepatitis A outbreak [71].

In this study, *aph(6)-Id/aph(3)-Ib/sul2/tet(A)* carriage was on a large (~103 kbp) IncB/O/K/Z plasmid with Australian links. *bla*<sub>CTX-M-27</sub>-positive *S. sonnei* has recently driven a prolonged MSM outbreak in Australia [17], with one successful sublineage harbouring an AMR profile consistent with one isolate in this study, 598080 (*mphA*, *ermB*, *dfrA1* and *sul2* genes with *gyrA* S83L). Typing of *S. sonnei* isolates 598080 and 607387 determined them as VN2.KH1.Aus; Australian isolates emerging from Kahnh Hoa Subclone 1 [47]. Though

not designated as MSM-linked by this scheme, this could represent travel-associated introduction into England and subsequent MSM transmission.

We found *bla*<sub>CTX-M-27</sub> was present on IncFII plasmids with high identity to a known plasmid linked to an English MDR *S. sonnei* cluster, p183660 (which harbours *bla*<sub>CTX-M-27</sub> and has high identity to pKSR100) [15]. We observed *bla*<sub>CTX-M-27</sub> displace *bla*<sub>TEM-1</sub>, and *bla*<sub>CTX-M-27</sub> loss between sequencing rounds was observed in *S. flexneri* 3a. This gene possibly has lower stability in this species; none of the 49 other *S. flexneri* isolates harboured *bla*<sub>CTX-M-27</sub> and globally *bla*<sub>CTX-M-27</sub> seems relatively rare in *S. flexneri* [72]. Alternatively, as this isolate also harboured chromosomal  $\beta$ -lactamases *ampC* and *bla*<sub>OXA-1</sub>, IS activity may have led to rapid plasmid *bla*<sub>CTX-M-27</sub> loss in the absence of antimicrobial pressure. It is interesting that isolate 607387 lost both *mphA* and *ermB* (two macrolide resistance mechanisms; phosphotransferase-mediated modification and methylase ribosome target-site modification) together with two MSM clade 2 members, whereas many *S. sonnei* isolates in non-MSM clades harboured only *mphA* (no isolate contained *ermB* alone).

Insertion sequences are challenging to identify with short-read data but facilitate ongoing species diversification in *Shigella* through gene interruption, deletion and genome reorganisation with convergence on streamlined genomes [73]. Hawkey et al. showed expansion of IS1, IS2, IS4, IS600 and IS911 in *Shigella* genomes; similarly, we showed high proportions of IS1, IS2 and IS4 with differing *S. sonnei* and *S. flexneri* IS profiles. IS activity causes potentially deleterious outcomes (e.g. ARG loss) but allows flexibility to create diverse structures; some of which are likely to be selected for [74]. It has been suggested if insertion of IS26 causes no deleterious effects it acts as a 'founder element' where further IS26 insertion occurs preferentially next to another [75], enabling rapid novel ARG acquisition. Concordantly, here IS26 preferentially inserted into IncFII plasmids and IS15 was present close to *bla*<sub>CTX-M-27</sub>, corresponding to one IS26 copy inserted into another. Similar to our genetic platform, *bla*<sub>CTX-M-27</sub> was flanked by IS26 and IS903B in *E. coli* ST38 IncF plasmids [76], suggesting a possible CTX-M reservoir. Currently annotated *bla*<sub>CTX-M-27</sub>-carrying *Shigella* plasmids are limited to *S. sonnei* p183660 (GenBank KX008967.1) and Swiss pEC732\_2 (GenBank CP015140.1). Within the latter a contrasting IS26-IS903B-*bla*<sub>CTX-M-27</sub>-ISEcp1 $\Delta$ -IS26 unit was identified [19], though it is possible IS26 was inserted into ISEcp1 (promoting high-level expression) in our study but not identified.

Although the mechanistic basis of *bla*<sub>CTX-M-27</sub> and *mphA*-*ermB* loss is speculative, this could reflect intramolecular IS26 transposition in cis, resulting in deletion of DNA between the IS and target site, leaving one IS copy [77]. IS26 is known to generate 8 bp target site duplications during cointegrate formation [77]; investigating these within additional ESBL plasmids over time would provide insights regarding temporal and spatial dynamics of AMR acquisition and loss.

This study is limited in low isolate numbers from a single region without patient exposure data or investigation of other CTX-M genes in additional isolates in PHE's collection; we cannot extrapolate ARG patterns and their position within MGEs in MSM-related strains over time or characterise risk factors. In *S. sonnei* there is generally high concordance between WGS ARG presence and phenotype when 100% read coverage and >90% nucleotide identity thresholds are used [69], though phenotypic data did not correlate with genotype for tetracycline resistance in 893916 and the underlying mechanism was not investigated. If long-read sequencing becomes viable on a routine scale, this would significantly aid plasmid monitoring to predict AMR dissemination, especially as *bla*<sub>CTX-M-27</sub> seems to have been acquired on multiple occasions.

## CONCLUSIONS

*Shigella* spp. are concerning due to their extraordinary ability to acquire and disseminate AMR. As ESBL genes can be propagated by movement on small mobile genetic elements, dissemination of 'epidemic' plasmids and via clonal spread, WGS surveillance of all three avenues with international data sharing is necessary to inform coordinated and responsible antimicrobial strategies.

### Funding information

D.R.G., C.J. and T.J.D., are affiliated to the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections at University of Liverpool in partnership with Public Health England (PHE), in collaboration with University of Warwick, and are based at PHE. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care or Public Health England.

### Acknowledgements

Special thanks to Matt Bird for his ongoing encouragement.

### Author contributions

Conceptualisation and methodology: L.A.C., D.R.G., C.J., T.J.D., Supervision: L.A.C. Investigation: R.K.L., D.R.G., Visualisation: R.K.L., Writing – original draft: R.K.L., Writing – review and editing: R.K.L., L.A.C., C.J., D.R.G.

### Conflicts of interest

The authors declare that there are no conflicts of interest.

### References

1. 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.
2. DuPont HL, Levine MM, Hornick RB, Formal SB. Inoculum size in shigellosis and implications for expected mode of transmission. *J Infect Dis* 1989;159:1126–1128.
3. Murray K, Reddy V, Kornblum JS, Waechter H, Chicaiza LF, et al. Increasing antibiotic resistance in *Shigella* spp. From infected New York city residents, New York, USA. *Emerg Infect Dis* 2017;23:332–335.
4. Schroeder GN, Hilbi H. Molecular pathogenesis of *shigella* spp.: Controlling host cell signaling, invasion, and death by type iii secretion. *Clin Microbiol Rev* 2008;21:134–156.
5. Yang F, Yang J, Zhang X, Chen L, Jiang Y, et al. Genome dynamics and diversity of *Shigella* species, the etiologic agents of bacillary dysentery. *Nucleic Acids Res* 2005;33:6445–6458.

6. Bardhan P, Faruque ASG, Naheed A, Sack DA. Decrease in shigellosis-related deaths without *shigella* spp.- specific interventions, Asia. *Emerg Infect Dis* 2010;16:1718–1723.
7. 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. *The Lancet* 2013;382:209–222.
8. Baker KS, Dallman TJ, Field N, Childs T, Mitchell H, et al. Genomic epidemiology of *Shigella* in the United Kingdom shows transmission of pathogen sublineages and determinants of antimicrobial resistance. *Scientific Reports* 2018;8.
9. Simms I, Field N, Jenkins C, Childs T, Gilbert VL, et al. Intensified shigellosis epidemic associated with sexual transmission in men who have sex with men - *Shigella flexneri* and s. *Sonnei* in England, 2004 to end of February 2015. *Eurosurveillance* 2015;20:1–5.
10. Williams PCM, Berkley JA. Guidelines for the treatment of dysentery (*shigellosis*): a systematic review of the evidence. *Paediatrics and International Child Health* 2018;38.
11. 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.
12. Rajakumar K, Bulach D, Davies J, Ambrose L, Sasakawa C, et al. Identification of a chromosomal *Shigella flexneri* multi-antibiotic resistance locus which shares sequence and organizational similarity with the resistance region of the plasmid NR1. *Plasmid* 1997;37:159–168.
13. Ye C, Lan R, Xia S, Zhang J, Sun Q, et al. Emergence of a new multidrug-resistant serotype X variant in an epidemic clone of *Shigella flexneri*. *J Clin Microbiol* 2010;48:419–426.
14. Chung The H, Boinett C, Pham Thanh D, Jenkins C, Weill F-X, et al. Dissecting the molecular evolution of fluoroquinolone-resistant *Shigella sonnei*. *Nat Commun* 2019;10:4828.
15. Baker KS, Dallman TJ, Field N, Childs T, Mitchell H, et al. Horizontal antimicrobial resistance transfer drives epidemics of multiple *Shigella* species. *Nature Communications* 2018;9:1–10.
16. Mook P, McCormick J, Bains M, Cowley LA, Chattaway MA, et al. ESBL-Producing and macrolide-resistant *Shigella sonnei* infections among men who have sex with men, England, 2015. *Emerg Infect Dis* 2016;22:1948–1952.
17. Ingle DJ, Andersson P, Valcanis M, Barnden J, Gonçalves da Silva A, et al. Prolonged outbreak of multidrug-resistant *Shigella sonnei* harbouring bla CTX-M-27 in Victoria, Australia. *Antimicrob Agents Chemother* 2020;64.
18. Muthurulandi Sethuvel DP, Veeraraghavan B, Vasudevan K, Devanga Ragupathi NK, Murugan D, et al. Complete genome analysis of clinical *Shigella* strains reveals plasmid pSS1653 with resistance determinants: A triumph of hybrid approach. *Gut Pathog* 2019;11:55.
19. Campos-Madueno EI, Bernasconi OJ, Moser AI, Keller PM, Luzzaro F, et al. Rapid Increase of CTX-M-Producing *Shigella sonnei* Isolates in Switzerland: spread of common plasmids and International Clones. *Antimicrobial Agents and Chemotherapy* 2020;01057–20.
20. Arredondo-Alonso S, Willems RJ, van Schaik W, Schürch AC. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microbial Genomics* 2017;3:10.
21. George S, Pankhurst L, Hubbard A, Votintseva A, Stoesser N, et al. Resolving plasmid structures in enterobacteriaceae using the Minion nanopore sequencer: Assessment of Minion and Minion/Illumina hybrid data assembly approaches. *Microb Genom* 2017;3:e000118.
22. Sedlazeck FJ, Lee H, Darby CA, Schatz MC. Piercing the dark matter: Bioinformatics of long-range sequencing and mapping. *Nat Rev Genet* 2018;19:329–346.
23. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–2120.
24. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 2018;34:2666–2669.
25. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–546.
26. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
27. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009;25:1754–1760.
28. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.
29. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE* 2014;9:e112963.
30. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res* 2017;27:737–746.
31. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
32. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, et al. Circlator: Automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 2015;16:294.
33. Darling AE, Mau B, Perna NT. Progressivmauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* 2010;5:e11147.
34. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
35. Arredondo-Alonso S, Rogers MRC, Braat JC, Verschuuren TD, Top J, et al. mlplasmids: a user-friendly tool to predict plasmid- and chromosome-derived sequences for single species. *Microb Genom* 2018;4:e000224.
36. Larsen M, Cosentino S, Rasmussen S, Friis C, Hasman H, et al. Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* 2012;50:1355–1361.
37. Varani AM, Siguier P, Goubeyre E, Charneau V, Chandler M. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol* 2011;12:R30.
38. Carattoli A, Hasman H. PlasmidFinder and *In silico* pMLST: Identification and typing of plasmid replicons in whole-genome sequencing (WGS). In: *Methods in Molecular Biology*. Humana Press Inc, 2020. pp. 285–294.
39. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, et al. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012;67:2640–2644.
40. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, et al. CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research* 2019;48:D517–D525.
41. Li H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–3100.
42. Milne I, Bayer M, Cardle L, Shaw P, Stephen G, et al. Tablet-next generation sequence assembly visualization. *Bioinformatics* 2009;26:401–402.
43. Xie Z, Tang H. ISEScan: Automated Identification of Insertion Sequence Elements in Prokaryotic Genomes. *Bioinformatics* (Oxford, England), 2017, pp. 3340–3347.
44. Dallman T, Ashton P, Schafer U, Jironkin A, Painset A, et al. SnaperDB: a database solution for routine sequencing analysis of bacterial isolates. *Bioinformatics* 2018;34:3028–3029.
45. Bardsley M, Jenkins C, Mitchell HD, Mikhail AFW, Baker KS, et al. Persistent transmission of *shigellosis* in England is associated with a recently emerged multidrug-resistant strain of *Shigella sonnei*. *J Clin Microbiol* 2020;58:3.



46. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 2020;37:1530–1534.
47. Hawkey J, Paranagama K, Baker KS, Bengtsson RJ, Weill FX, et al. Global population structure and genotyping framework for genomic surveillance of the major dysentery pathogen, *Shigella sonnei*. *Nature Communications* 2021;12:1–12.
48. Hunt M, Bradley P, Lapierre SG, Heys S, Thomsen M, et al. Antibiotic resistance prediction for *Mycobacterium tuberculosis* from genome sequence data with Mykrobe. *Wellcome Open Res* 2019;4:191.
49. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–359.
50. Sullivan MJ, Petty NK, Beatson SA. Easyfig: A genome comparison visualizer. *Bioinformatics* 2011;27:1009–1010.
51. Petkau A, Stuart-Edwards M, Stothard P, van Domselaar G. Interactive microbial genome visualization with GView. *Bioinformatics* 2010;26:3125–3126.
52. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genomics* 2011;12:402.
53. Yu G. Using ggtree to Visualize Data on Tree-Like Structures. *Current Protocols in Bioinformatics* 2020.
54. Buchrieser C, Glaser P, Rusniok C, Nedjari H, D'Hauteville H, et al. The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. *Mol Microbiol* 2000;38:760–771.
55. Faherty C, Harper JM, Shea-Donohue T, Barry EM, Kaper JB, et al. Chromosomal and plasmid-encoded factors of *Shigella flexneri* induce secretogenic activity ex vivo. *PLoS ONE* 2012;7:49980.
56. Mattock E, Blocker AJ. How do the virulence factors of shigella work together to cause disease. *Front Cell Infect Microbiol* 2017;7:64.
57. Bagel S, Hüllen V, Wiedemann B, Heisig P. Impact of gyrA and parC mutations on quinolone resistance, doubling time, and supercoiling degree of *Escherichia coli*. *Antimicrob Agents Chemother* 1999;43:868–875.
58. Ranjbar R, Farahani A. Shigella: Antibiotic-resistance mechanisms and new horizons for treatment. *Infect Drug Resist* 2019;12:3137–3167.
59. Ingle DJ, Easton M, Valcanis M, Seemann T, Kwong JC, et al. Co-circulation of multidrug-resistant *Shigella* among men who have sex with men in Australia. *Clin Infect Dis* 2019;69:1535–1544.
60. Yousfi K, Gaudreau C, Pilon PA, Lefebvre B, Walker M, et al. Genetic mechanisms behind the spread of reduced susceptibility to azithromycin in *Shigella* strains isolated from men who have sex with men in Québec, Canada. *Antimicrob Agents Chemother* 2019;63.
61. Trieu-Cuot P, Labigne-Roussel A, Courvalin P. An IS15 insertion generates an eight-base-pair duplication of the target DNA. *Gene* 1983;24:125–129.
62. Klontz KC, Singh N. Treatment of drug-resistant *Shigella* infections. *Expert Rev Anti Infect Ther* 2015;13:69–80.
63. Zamanlou S, Ahangarzadeh Rezaee M, Aghazadeh M, Ghotaslou R, Babaie F, et al. Characterization of integrons, extended-spectrum  $\beta$ -lactamases, AmpC cephalosporinase, quinolone resistance, and molecular typing of *Shigella* spp. *Infect Dis (Lond)* 2018;50:616–624.
64. Medeiros PHQS, Lima AÂM, Guedes MM, Havt A, Bona MD, et al. Molecular characterization of virulence and antimicrobial resistance profile of *Shigella* species isolated from children with moderate to severe diarrhea in northeastern Brazil. *Diagn Microbiol Infect Dis* 2018;90:198–205.
65. Sethuvel DPM, Anandan S, Michael JS, Murugan D, Neeravi A, et al. Virulence gene profiles of *Shigella* species isolated from stool specimens in India: its association with clinical manifestation and antimicrobial resistance. *Pathog Glob Health* 2019;113:173–179.
66. Cheasty T, Day M, Threlfall EJ. Increasing incidence of resistance to nalidixic acid in shigellas from humans in England and Wales: Implications for therapy. *Clin Microbiol Infect* 2004;10:1033–1035.
67. The HC, 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.
68. Bevan E, Jones A, Hawkey P. Global epidemiology of CTX-M  $\beta$ -lactamases: temporal and geographical shifts in genotype. *J Antimicrob Chemother* 2017;72:2145–2155.
69. 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.
70. Gilbert VL, Simms I, Jenkins C, Furegato M, Gobin M, et al. Sex, drugs and smart phone applications: Findings from semistructured interviews with men who have sex with men diagnosed with *Shigella flexneri* 3a in England and Wales. *Sex Transm Infect* 2015;91:598–602.
71. Ruscher C, Werber D, Thoulas J, Zimmermann R, Eckardt M, et al. Dating apps and websites as tools to reach anonymous sexual contacts during an outbreak of hepatitis A among men who have sex with men, Berlin, 2017. *Eurosurveillance* 2019;24:1800460.
72. Liu G, Qian H, Tang B, Chen Y, Kang H, et al. Prevalence and characterisation of third-generation cephalosporin-resistant *Shigella flexneri* isolates from Jiangsu Province, China, 2013–2015. *J Glob Antimicrob Resist* 2018;15:283–287.
73. Hawkey J, Monk JM, Billman-Jacobe H, Palsson B, Holt KE. Impact of insertion sequences on convergent evolution of *Shigella* species. *PLoS Genet* 2020;16:e1008931.
74. Partridge SR. Analysis of antibiotic resistance regions in Gram-negative bacteria. *FEMS Microbiol Rev* 2011;35:820–855.
75. Harmer CJ, Moran RA, Hall RM. Movement of IS26-Associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26. *mBio* 2014;5:e01801–14.
76. Mostafa HH, Cameron A, Taffner SM, Wang J, Malek A, et al. Genomic surveillance of ceftriaxone-resistant *Escherichia coli* in Western New York suggests the extended-spectrum  $\beta$ -lactamase blaCTX-M-27 is emerging on distinct plasmids in ST38. *Front Microbiol* 2020;11:1747.
77. He S, Hickman AB, Varani AM, Siguier P, Chandler M, et al. Insertion sequence IS26 reorganizes plasmids in clinically isolated multidrug-resistant bacteria by replicative transposition. *mBio* 2015;6.