

Drivers of antimicrobial resistance amongst intestinal *Escherichia coli* isolated from children in South Asia and sub-Saharan Africa

Danielle J. Ingle^{1,2,#}, Myron M. Levine³, Karen L. Kotloff³, Kathryn E. Holt^{2†}
Roy M. Robins-Browne^{1,4,*†}

Short title: Antimicrobial resistance in enteropathogenic *Escherichia coli*

¹ Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, The University of Melbourne, Victoria, Australia

² Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Victoria, Australia

³ Departments of Pediatrics and Medicine, Center for Vaccine Development, University of Maryland School of Medicine, Baltimore, MD, USA

⁴ Murdoch Childrens Research Institute, Royal Children's Hospital, Parkville, Victoria, Australia

Current address: National Centre for Epidemiology & Population Health, The Australian National University, and Microbiological Diagnostic Unit Public Health Laboratory, Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, The University of Melbourne, Victoria, Australia

* Corresponding author:

E-mail: r.browne@unimelb.edu.au (RMRB)

† These authors contributed equally to this work.

Abstract

Antimicrobial resistance (AMR) dynamics are poorly understood in developing countries, where data on the prevalence of AMR in enteric bacteria are sparse, particularly among children and in the community setting. Here we use a combination of phenotyping, genomics and antimicrobial usage data to investigate patterns of AMR amongst atypical enteropathogenic *E. coli* (aEPEC) strains isolated from children <5 years old in seven countries (four in sub-Saharan Africa and three in South Asia) over a three-year period. We detected very high rates of AMR, with 65% of isolates displaying resistance to ≥ 3 drug classes; the rates of AMR were the same amongst strains associated with diarrhea and strains that were carried asymptotically. Whole genome sequencing identified a diversity of genetic mechanisms for AMR, which could explain >95% of observed phenotypic resistance. Analysis of AMR gene co-occurrence revealed clusters of acquired AMR genes that were frequently co-located on small plasmids and transposons, providing opportunities for acquisition of multidrug resistance in a single step. We used discriminant analysis to investigate potential drivers of AMR within the bacterial population, and found that genetic determinants of AMR were associated with geographical location of isolation but not with phylogenetic lineage of the *E. coli* strain or disease status of the human host. Comparison with antimicrobial usage data showed that the prevalence of resistance to newer drugs (fluoroquinolones and third-generation cephalosporins) was correlated with usage, which was generally higher in South Asia than Africa. In particular, fluoroquinolone resistance-associated mutations in *gyrA* were significantly associated with use of these drugs for treatment of diarrheic children. Notably resistance to older drugs such as trimethoprim, chloramphenicol and ampicillin, which are conferred by acquired AMR genes that were frequently clustered together in mobile genetic elements, were common in all locations despite differences in usage; this suggests that reversion to sensitivity is unlikely to occur even if these drugs are removed from circulation. This study provides much-needed insights into the frequencies of AMR in intestinal *E. coli* in community-based children in developing countries and to antimicrobial usage for diarrhea where the burden of infections is greatest.

Introduction

Certain pathotypes of *Escherichia coli* (*E. coli*) are a leading cause of diarrhea in children, especially in the developing countries of sub-Saharan Africa and South Asia¹. Pathogenic varieties of *E. coli* have evolved on multiple occasions; atypical enteropathogenic *E. coli* (aEPEC) is one such pathotype, which is defined by the presence of the locus of enterocyte effacement (LEE) pathogenicity island, and the absence of Shiga toxins (denoting enterohemorrhagic *E. coli*) and type IV bundle-forming pili (denoting typical EPEC)²⁻⁵. We have recently identified distinct lineages of aEPEC and ten common clonal groups (CGs)⁶. Importantly, unlike most other *E. coli* pathotypes, aEPEC has various animal reservoirs and, while it can cause a range of disease symptoms in humans from sporadic to persistent diarrhea, it may also colonize the human gut asymptomatically^{1,7-9}.

Antimicrobial resistance (AMR) has been reported in *E. coli* from multiple animal sources, the environment as well as in hospital-acquired infections globally¹⁰⁻¹⁵. Of greatest concern are multi-drug resistant (MDR) *E. coli* (non-susceptible to one or more agents in at least three antimicrobial categories)¹⁶, particularly those that are fluoroquinolone resistant and/or produce extended-spectrum beta-lactamases (ESBL). While several recent studies have reported increases in ESBLs^{17,18}, gentamicin^{19,20}, ciprofloxacin²⁰⁻²⁴ and revealed diverse AMR profiles, these data have largely come from *E. coli* responsible for disease in hospital settings. Thus, there remain significant gaps in knowledge of the prevalence of AMR in human intestinal *E. coli* populations globally, particularly in the developing nations of sub-Saharan Africa and South Asia²⁵.

The rise of AMR, and particularly MDR in bacteria, is a significant public health concern as it makes infections non-responsive to treatment while increasing the reservoir of resistance-encoding genes. Although antimicrobials are not recommended for the treatment of uncomplicated gastroenteritis, they are commonly administered to diarrheic children in developing countries to prevent or treat dysentery²⁶ and prolonged diarrhea, of which aEPEC is a major cause^{8,9}. Enhancing our knowledge of AMR amongst gut-dwelling *E. coli* has broader implications: first, because *E. coli* is a leading cause of extra-intestinal infections in hospitals, and strains colonizing the gastrointestinal tract of patients are the major reservoir of these infections; and second, because most AMR in *E. coli* is encoded on mobile genetic elements that are subject to HGT, which enable the rapid dissemination of genes that prevent susceptibility to drugs and will be maintained under positive selective pressure^{27,28}. Hence AMR determinants may be shared between *E. coli* strains regardless of pathotype. For these reasons, it is useful to determine the AMR profiles of intestinal *E. coli*, such as aEPEC, as these can provide an indication of the prevalence of AMR in Enterobacteriaceae in general and the gut microbiome as a whole.

Here we present AMR data for 185 aEPEC isolates collected as part of the Global Enteric Multicenter Study (GEMS)^{1,29}, using phenotypic susceptibility data and whole genome sequence analysis to identify the mechanisms of resistance and potential drivers of variation in AMR profiles. These isolates, collected from healthy children in the community and children with diarrhea at seven sites in sub-Saharan Africa and South Asia, provide a unique opportunity to investigate the prevalence of AMR in a bacterial population that was not selected on the basis of AMR profile.

Results

Antimicrobial susceptibility profiles

Our collection of 185 aEPEC isolates (Table S1) was tested for susceptibility to 16 antimicrobials (Table S2). Resistance was detected to 14 of the drugs investigated (Fig 1A; Table S3), and 121 isolates (65%) were MDR (i.e., resistant to ≥ 3 drug classes) (Fig 1B). No resistance was detected to amikacin (an aminoglycoside) or meropenem (a carbapenem). Only 35 isolates (19%) were susceptible to all drugs tested (17 from cases and 18 from controls). Resistance to 'older' antimicrobials was common, with 121 (65%) of isolates resistant to ampicillin, 80 (43%) resistant to streptomycin, 124 (67%) resistant to trimethoprim, 122 (66%) resistant to

trimethoprim/sulfamethoxazole and 104 (56%) resistant to tetracycline (Fig 1A). Around half (n=96, 52%) of isolates were resistant to three or more of these first-line drugs. While streptomycin resistance was common, 180 isolates (97%) were susceptible to all other aminoglycosides tested; the exceptions being one isolate from India that was resistant to tobramycin and intermediately susceptible to gentamicin; and four isolates from India that were resistant to both drugs. Fluoroquinolone resistance was relatively infrequent, with 31 (17%) isolates resistant to norfloxacin and 8 (4%) resistant to both norfloxacin and ciprofloxacin. Resistance to chloramphenicol (11%) and azithromycin (7%) was also infrequent. Amongst the beta-lactam antibiotics, while ampicillin resistance was common (65%), resistance to the cephalosporins, ceftriaxone (third generation; 3%), ceftazidime (third generation; 3%) and cefepime (fourth generation; 1.6%) was rare (Fig 1A-B).

Genetic determinants of AMR

The genomes of the 185 aEPEC isolates were screened for known genetic determinants of AMR, including acquired genes as well as point mutations in chromosomal genes associated with resistance to fluoroquinolones and nitrofurantoin (Fig 1C; Fig S1 and Table S4). Thirty acquired AMR genes were detected, along with four point mutations (two in *gyrA*, one in *parC*, and one in *nfsA*). Four acquired AMR genes were detected in more than half the isolates: *bla*_{TEM} genes (ampicillin), *strA* and *strB* (streptomycin), and *sul2* (sulfonamides). Alleles of dihydrofolate reductase (*dfr*) genes encoding trimethoprim resistance were detected in 132 (71%) isolates. The most common of these were *dfrA1* (12%), *dfrA5* (23%), *dfrA7* (20%), and *dfrA8* (16%). The common AMR genes were determined to be predominately co-mobilized on a small number of common mobile elements (Fig S2; Fig S3 and Supplementary Results). Extensive diversity of AMR genotypes was observed, with 104 distinct combinations of AMR determinants across the 185 isolates (Fig S1).

Prediction of AMR phenotypes from genotypes

Phenotypic resistance profiles were largely explained by known genetic determinants of AMR (Fig 1, Table 1). For most drugs, the detection of known genetic determinants was both sensitive (>95%) and specific (>90%) for AMR phenotypes. The frequency of very major errors (failing to detect resistance when it is present) exceeded the minimum acceptable threshold of 1.5% for five drugs (Table 1). These were: ampicillin (4.9%); streptomycin (2.2%); trimethoprim (2.2%); trimethoprim/sulfamethoxazole (2.2%) and tetracycline (2.2%). Major errors (predicting resistance when none is present) were also detected for these and several other antimicrobials (Table 1). The highest major error rates were observed for streptomycin (26.5%), ampicillin (7.0%), trimethoprim (5.4%), and tetracycline (4.3%). The sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for each drug are summarized in Table S5. Sensitivity and NPV were greater than 90% for all drugs tested with the exception of azithromycin (85% sensitivity) and ampicillin (85% NPV). Specificity and PPV were more variable, reflecting the error rates outlined above (Table 1).

Table 1. Comparison of phenotypic and genotypic AMR profiles of 185 aEPEC

	Resistant phenotype	Resistant genotype	False susceptible [^]	Susceptible phenotype	Susceptible genotype	False resistant ^{^^}
Beta-lactams						
Ampicillin	121	112	9 (4.9%)	64	51	13 (7.0%)
Cefepime	3	3	0	182	178	4 (2.2%)
Ceftriaxone	6	6	0	179	174	5 (2.7%)
Ceftazidime	6	6	0	179	174	5 (2.7%)
Meropenem	0	0	0	185	185	0
Aminoglycosides						
Streptomycin	80	76	4 (2.2%)	105	56	49 (26.5%)
Gentamicin	5	5	0	180	180	0
Tobramycin	5	5	0	180	180	0
Amikacin	0	0	0	185	185	0
Folate pathway inhibitors						
Trimethoprim	124	120	4 (2.2%)	61	50	11 (0.9%)
Trimethoprim/sulfamethoxazole*	122	118	4 (2.2%)	63	53	10 (5.4%)
Nitrofurantoin						
Nitrofurantoin	0	0	0	185	182	3 (1.6%)
Chloramphenicol						
Chloramphenicol	21	19	2 (1.1%)	164	160	4 (2.2%)
Macrolides						
Azithromycin	13	11	2 (1.1%)	172	166	6 (3.2%)
Tetracyclines						
Tetracycline	104	100	4 (2.2%)	81	73	8 (4.3%)
Fluoroquinolones						
Ciprofloxacin**	8	8	0	177	173	4 (2.2%)
Norfloxacin	31	30	1 (0.5%)	154	149	5 (2.7%)

* When two genes are required for resistance, both were required for genotypic resistance: *strA* and *strB* for streptomycin, and a *dfr* gene plus a *sul* gene for resistance to trimethoprim/sulfamethoxazole

**At least one point mutation in *gyrA* and a second in *gyrA*, *gyrB*, *parC* or *parE*

[^] very major errors of >1.5% shown in boldface font

^{^^} major errors of >3.0% shown in boldface font

Sources of variation in AMR profiles

Given the diversity of AMR profiles observed in our aEPEC collection, we sought to determine if the distribution of AMR genes was associated with disease status, phylogenetic lineage, or geographic location where the strain was isolated. First we compared the frequency of AMR phenotypes and genotypes amongst aEPEC isolated from diarrhea cases and asymptomatic controls. Only data from confirmed cases (n=94) and controls (n=88) were used for this analysis. For each drug class, neither AMR phenotype, nor AMR predicted from genotype, was statistically different between cases and controls (Table S6). Also, the frequencies of individual AMR genes in isolates from case and controls were similar (Fig 2A). As AMR determinants were equally distributed amongst cases and controls, all isolates were pooled together for further analysis of lineage and region.

Discriminant analysis of principal components (DAPC)³⁰ on the binary matrix of AMR genetic determinants (Table S4; Fig 3B-C) revealed that the first 15 PCs accounted for >93% of the variance in AMR profiles, and were retained for discriminant analysis by phylogenetic lineage (clonal groups, defined in Ingle *et al.*⁶ and highlighted in Fig S1) or by region of origin (East Africa, West Africa and Asia; see Fig 2 and Fig S4)).

Variation in AMR determinants were not associated with clonal group, except for CG378 which was characterized by a lack of most of the common AMR genes (*bla*_{TEM} variants, *sul2*, *tet(A)* and *tetB*) and the presence of the rarer *catA* gene (present in 7 of 9 CG378 isolates compared to 12.5% of non-CG378 isolates; OR=24, $P < 10^{-4}$). Variation in AMR determinants was associated with region of origin (Fig 2B, Fig S4B). Discriminant function 1 (DF1) separated Asian from African isolates and was associated with *gyrA* SNPs; DF2 separated East from West African isolates and was associated with *dfrA8*, *dfrA5*, *tet(A)* and *sul2*. The various *dfrA* genes were differentially distributed across the various GEMS sites (Fig 2D). For example, *dfrA1* predominated at Asian sites and *dfrA8* was commonest at West African sites whilst *dfrA14* and *dhfr7* were common in Mozambique and Kenya, respectively (Fig 2D). Further, *tet(A)* was more common at West and East African sites than in Asia. These genetic differences were reflected in AMR phenotypes, as resistance to ciprofloxacin and third-generation cephalosporins was identified only in strains from Asia, while resistance to tetracycline was more common in African than in Asian isolates (Fig 3A).

Differences in local antimicrobial drug usage

We hypothesized that regional differences in AMR and associated genetic determinants may reflect differences in local selective pressure from antimicrobial exposure. The broad patterns of antimicrobial use for the treatment of diarrhea across the study sites showed that trimethoprim/sulfamethoxazole and penicillins were used more frequently at African sites, while macrolides (azithromycin) and fluoroquinolones (in particular ciprofloxacin) were used more frequently in Asia (Fig 3B-C). These patterns of antimicrobial usage showed some association with AMR phenotypes insofar as we observed higher levels of both usage and resistance for azithromycin and fluoroquinolones at Asian sites, and for trimethoprim at East African sites (Fig 3). We could not formally test these associations, however, due to the small numbers of observations at most regional subgroups and the limited variation in usage of most drugs across sites. We therefore investigated the associations between usage and resistance for the two drugs that showed substantial (>10%) usage at three or more study sites; these were ciprofloxacin and trimethoprim. Across the seven sites, ciprofloxacin usage was significantly associated with the prevalence of substitutions in the QRDR of *gyrA* and *parC* ($R^2 = 0.87$, $P = 0.002$; Fig 4). By contrast, trimethoprim usage was not associated with the prevalence of horizontally acquired *dfr* genes that confer resistance to the drug ($R^2=0.04$, $P > 0.5$).

Discussion

This is the first study to report the prevalence of AMR in community-associated *E. coli* isolates at the GEMS sites. While limited AMR data for *E. coli* are available from the seven countries where the GEMS sites were located, there are no national surveillance data from any of these countries. Furthermore, AMR data on *E. coli* in these countries predominately pertain to *E. coli* causing extra-intestinal infections and to a limited number of drugs, mainly third-generation cephalosporins and fluoroquinolones^{21,23-25,31}. Here we present the first comprehensive AMR phenotypic data and molecular characterization for multiple antimicrobials, detecting several common mobile elements mediating the AMR including small plasmids, class I integrons with different cassettes and transposons including Tn6029, Tn10 and Tn1721 (Fig S3). Many of the common mobile elements we detected have been reported widely in *E. coli* and other Enterobacteriaceae from human intestinal and extra-intestinal and animal samples³²⁻³⁶. Thus, while not conducted for the specific purpose of AMR surveillance, these aEPEC are likely represent a holistic overview of *E. coli*, and possibly other Enterobacteriaceae, in circulation at the study sites. This suggestion is supported by the observation that in this study aEPEC from cases and asymptomatic controls exhibited the same AMR profiles.

The majority (65%) of aEPEC analyzed in this study were MDR, with many displaying resistance to three or more older drugs in key groups: beta-lactams (ampicillin), aminoglycosides (streptomycin), trimethoprim-sulfamethoxazole and tetracycline (Fig 1). The rates of resistance to these drugs generally were lower amongst Asian isolates (21-62%), whereas resistance to newer drugs such as ceftriaxone, fluoroquinolones and azithromycin were detected only at the Asian sites (Fig 3A). These patterns were broadly consistent with antimicrobial usage data from the study sites, which showed that ciprofloxacin and azithromycin were commonly used to treat diarrhea at the Asian sites, while trimethoprim-sulfamethoxazole was the mainstay of treatment at the African sites, supplemented occasionally by other older drugs including ampicillin and chloramphenicol (Fig 3B).

The high AMR frequencies for first-line drugs, in particular penicillins, detected in this study are similar to reports of other intestinal, extra-intestinal and environmental isolates of *E. coli*^{22,37}. Of note, the high levels of ciprofloxacin-resistance found in this study at the Asian sites corresponded to high levels reported in clinical cases in India, Bangladesh and Pakistan^{17,18,21,23,24,38}. Further, while ESBL-producing aEPEC were detected in this study in Asia, the frequency of these was lower than those reported from hospitals in both Asia and Africa^{18,21,31,38}. These findings suggest the prevalence of ESBLs, and indeed resistance to carbapenems and last-line aminoglycosides (which were not detected in these community aEPEC) are in higher circulation in hospital settings and have not yet become established in the community settings we investigated.

In agreement with our data relating to AMR phenotypes, we found that the genetic determinants of resistance were similar in bacteria isolated from diarrheal cases and asymptomatic controls (Fig 2A, Table S6) and were not associated with clonal background of the strain, but were associated with the geographic region where the bacteria were isolated (Fig 2B-C; Fig S4). These findings are consistent with the hypothesis that there is selection for different AMR-encoding genes in different locations due to the differing patterns of antimicrobial usage (Fig 3B-C). In strong support of this explanation, mutations in the quinolone resistance determining regions of *gyrA* and *parC* contributed to the regional discrimination function (Fig 2B) and were significantly associated with the frequency of ciprofloxacin use across the seven study sites (Fig 4). However, the situation was more complex for horizontally transferred AMR genes associated with resistance to older drugs. While individual *dfr* alleles were distributed differently across sites (Fig 2C) and contributed to orthogonal components of the regional discriminant function (Fig 2B), the overall prevalence of *dfr* genes was relatively high (50-90%) at each site and not significantly associated with use of trimethoprim (Fig 3D). Similarly, genes encoding resistance to ampicillin, streptomycin and tetracycline were also common in sites where these drugs were rarely used. This could be due to (i) a lack of fitness cost associated with these older resistances, resulting in maintenance of the genes in the *E. coli* population after drug usage declines^{11,13,39-41}; (ii) co-selection for resistance to old and new drugs whose associated genes are present in the same mobile elements^{42,43}; and/or (iii)

selection from drug exposure unrelated to the treatment of diarrhea; for example, the use of trimethoprim to treat other types of infection, or in animal or environmental reservoirs amongst which intestinal *E. coli* circulates. We could not distinguish between hypotheses (i) and (ii) based on our data reported here as we could not determine the precise location of AMR genes from the available short-read sequence data, although further experiments such as conjugation or long-read sequencing could resolve this in future. It is notable, however, that the pCERC-like plasmids (which often carried resistance to streptomycin, sulfonamides and trimethoprim), are very common in *E. coli*, possibly because their small size (~6 kbp, of which half encodes AMR genes) imposes a low fitness cost^{32,44}. It is also notable that these and many of the other AMR genes we detected are also associated with composite transposons which can integrate into the bacterial chromosome, where they are maintained at lower fitness cost than large AMR plasmids^{36,45,46}.

This study has several key strengths. First, the bacterial samples and usage data were collected seven study sites, four in Africa and three in south Asia, using the same protocols enabling reliable comparisons between the sites^{29,47}. Second, the bacterial samples include those from children with diarrheal symptoms and asymptomatic controls in the same study populations, enabling the inference that the aEPEC in this study are broadly representative of the *E. coli* in circulation at these sites. Third, the comprehensive characterization of AMR phenotypes to 16 antimicrobials *in vitro* and their molecular mechanisms from whole genome sequence enabled comparison of both approaches, and demonstrated the value of both types of data. In particular, it showed that reliance on AMR genetic determinants alone may be misleading due to occasional systematic errors in predicting AMR phenotype. Finally, we were able to investigate the complex relationship between antimicrobial use and resistance in community-associated fecal *E. coli* in a way that would be unfeasible using more limited techniques or study sites.

Despite these strengths, there were limitations to this study that could be areas for future improvement. First, the antimicrobial usage data collected at the GEMS sites captured only the drugs prescribed at the GEMS clinics for treatment of diarrheal disease, and does not include antimicrobials available from other sources in the study populations that could also select for AMR *E. coli*, including the treatment of respiratory tract infections³⁸. Each GEMS site was expected to follow the WHO guidelines for the management of diarrheal disease¹, but non-prescribed antimicrobials may account for 19-100% of antimicrobial use outside northern Europe and North America⁴⁸. Increased surveillance and monitoring of antimicrobial use for not only diarrheal disease, but other bacterial infections and their use in agriculture, should help to address this limitation in future. Second, a limitation of the genomic analysis was that the short-read data were insufficient to resolve how all of the AMR-associated genes were mobilized in the aEPEC genomes, which is a common issue in high-throughput genomics studies in general. This could be resolved using long-read sequencing^{36,49}, which is becoming more affordable to an extent where it may be feasible to resolve AMR fully in hundreds of bacterial genomes⁵⁰.

Conclusions

Our data fill important information gaps concerning the prevalence of AMR amongst intestinal *E. coli* circulating in children in seven developing countries in Africa and Asia. In particular, the study showed that MDR involving resistance to multiple first-line or ‘older’ drugs (ampicillin, tetracycline, streptomycin and trimethoprim-sulfamethoxazole) was common at all sites, and resistance to the ‘newer’ antimicrobials, such as fluoroquinolones and azithromycin, had emerged only in Asian sites where these drugs were being used for the treatment of diarrhea. Resistance to older drugs was also common at these sites, such that only Asian isolates were resistant to seven or eight categories of antimicrobials, indicating that changing patterns of antimicrobial use leads to an accumulation of resistance determinants rather than their replacement. Importantly, we found that many AMR genes were encoded on mobile elements that circulate widely amongst *E. coli* and other Enterobacteriaceae, suggesting that these findings may be generalizable to other enteric pathogens and members of the gut microbiome.

Methods

aEPEC isolates and corresponding whole genome sequences

A total of 185 aEPEC isolates from the GEMS culture collection were included in the analysis^{1,6,29}. Their collection, whole genome sequencing and phylogenomic analysis have been described previously⁶. Briefly, the bacteria were isolated in pure culture from fecal samples from children aged 0-5 years with (cases, n = 94) or without (controls, n = 88) diarrhea at study sites located in seven developing countries (The Gambia, Mali, Kenya, Mozambique, Bangladesh, India and Pakistan)⁴⁷. Three isolates were from children whose case/control status was uncertain. Fecal samples were collected at the study sites before antimicrobial treatment, although prior exposure to antimicrobials from other sources cannot be ruled out. Control children were also not receiving any antimicrobial treatment.

Whole genome sequences were generated for all 185 aEPEC isolates via the Illumina HiSeq platform (100 bp paired-end reads) and assembled using Velvet, as described previously^{6,51}. Details of the individual isolates, accessions numbers for the corresponding genome sequence reads and assemblies (deposited collectively under BioProject ERP001141), and associated metadata are provided in Table S1.

Phenotypic characterization of antimicrobial resistance profiles

Antimicrobial susceptibility testing (AST) to 16 antimicrobials was performed using the VITEK2 (bioMérieux) system or an agar-dilution method. A summary of the drugs, testing methods, and the MIC breakpoints used to determine susceptible (S), intermediate (I) or resistant (R) status for each drug is shown in Table S2. The controls used were three *Salmonella enterica* isolates with known resistance profiles. These were E0001 *Salmonella* Typhimurium PT1 (susceptible to all drugs tested), E004 *Salmonella* Heidelberg (resistant to ampicillin, streptomycin, tetracycline, chloramphenicol, sulfathiazole, trimethoprim, kanamycin, spectinomycin and gentamicin) and E2060 *Salmonella* Thompson (resistant to streptomycin_{mod}, tetracycline, kanamycin, nalidixic acid and ciprofloxacin).

For VITEK2 assays, pure isolates were streaked on MacConkey agar plates and grown at 37°C overnight. Isolates were then sub-cultured onto horse blood agar (HBA) plates for fresh culture and grown overnight at 37°C. One to three colonies were selected from each HBA plate and suspended in saline to an OD of ~0.5 MacFarlane Units before being subjected to VITEK2 analysis.

Susceptibility to streptomycin, chloramphenicol, azithromycin, and tetracycline were determined using an agar-dilution method. Bacterial suspensions were first prepared as described above. To each of 32 stainless steel wells, 450 µL of nutrient broth containing 0.05% agar were added, followed by 50 µL of a bacterial suspension. Each Mueller Hinton (MH) agar antimicrobial-containing plate for susceptibility testing and two control MH and MacConkey agar plates were inoculated using a 32-pin replicator. Each pin delivers 2 µL to the plate such that the final number of CFU in each sample was ~10⁴. Plates were incubated overnight at 37°C and inspected the following day. Growth on an antimicrobial-containing MH plate was recorded as phenotypically resistant to the drug, whilst no growth was recorded as susceptible.

The European Committee on Antimicrobial Susceptibility Testing (EUCAST) MIC breakpoints were used where available⁵². Differences exist between the EUCAST and Clinical and Laboratory Standards Institute (CLSI) guidelines in terms of MIC breakpoints and the drugs to be tested. Specifically, tetracycline does not have defined MIC breakpoints under the EUCAST scheme. Hence the CLSI MIC breakpoint was used. Streptomycin and azithromycin do not have established MIC breakpoints under either scheme^{53,54}. Previous research proposed a breakpoint of 16 µg/mL for streptomycin in *E. coli*⁵⁴. Little information is available on the MIC distribution of azithromycin for *E. coli*. A breakpoint of 16 µg/mL has been proposed for *Salmonella enterica* based on a study in which the majority of isolates displayed MIC values of 4 to 8 µg/ml⁵³. For the

present study, the breakpoint for each of these drugs was set at the conservative MIC of 16 µg/mL. The antimicrobial susceptibility data for each isolate are shown in Table S3.

Detection of antimicrobial resistance genes

An SRST2-formatted version of the ARG-ANNOT AMR gene database⁵⁵ was downloaded from <https://github.com/katholt/srst2>. All sequence read sets were screened against the database using SRST2 to detect the presence of known acquired resistance genes in each genome⁵⁶. The beta-lactamase genes, *ampC1*, *ampC2*, *ampH*, were excluded from analysis, as in *E. coli* they are core genes that normally do not confer antibiotics resistance. The results were transformed into a binary table in R to indicate presence/absence of acquired resistance gene alleles (Table S4).

Detection of SNPS conferring resistance to fluoroquinolones and nitrofurantoin

Chromosomal mutations known to be associated with resistance to fluoroquinolones in *E. coli* were extracted from the genome-wide SNP calls obtained previously based on mapping the reads to the *E. coli* strain 12009 O103:H2 reference genome⁶. These included specific mutations in the QRDR of *gyrA*, *gyrB*, *parC* and *parE*⁵⁷, and non-synonymous substitutions in *nfsA* (residues 11-15) which confer resistance to nitrofurantoin^{58,59}.

Statistical analysis of AMR phenotype prediction from genotype

The ability of genotypes to predict drug susceptibility phenotypes was assessed by comparing summarized AMR phenotypes (S, I and R) with the presence of known AMR-associated genes and mutations. Antimicrobial susceptibility prediction errors are generally characterized as either very major (calling a resistant isolate susceptible) or major (calling a susceptible isolate resistant)^{60,61}. The current standard for acceptable very major error and major error rates are <1.5% and <3%, respectively⁶⁰. Here, very major errors were said to have occurred when an isolate was phenotypically resistant, but no known resistance genes or mutations were detected, while major errors were made when an isolate carried known resistance determinant(s) but was phenotypically susceptible. Statistical analysis to determine sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated in the epiR package for R, using the *epi.stats* function with 0.95 confidence intervals⁶² for each antimicrobial tested.

Genomic reconstruction of demographic groups by discriminant analysis of principal components

The binary matrix of AMR genetic determinants was used as the input for discriminant analysis of principal components (DAPC) implemented in the adegenet package in R³⁰. The first fifteen principal components (PCs), which together explained >95% of the variance in AMR gene content, were retained for discriminant analysis (DA) to explore the ability of PCs to discriminate between groups of strains defined by geographic region of origin (West Africa: The Gambia and Mali; East Africa: Kenya and Mozambique, and Asia: Bangladesh, India and Pakistan) or clonal group membership. The two PCs contributing the most to DA were plotted, and labeled with the genetic determinants whose variation contributed the most to those components. The posterior group membership probabilities for each discriminant function were also plotted.

Construction of co-occurrence network

A pairwise co-occurrence matrix of acquired AMR genes was constructed by transforming the binary AMR gene content matrix (Table S3) in R. The co-occurrence relationships were visualized between all pairs of genes using heatmap (Fig S1). Networks of co-occurring genes, in which nodes represent genes and edges represent a frequency of co-occurrence exceeding a given threshold (set to ≥ 20 , ≥ 33 , ≥ 46 , ≥ 100 genomes), were visualized in R using the igraph package⁶³.

Plasmid replicon screening

An SRST2-formatted version of the PlasmidFinder database⁶⁴ was downloaded from

<https://github.com/katholt/srst2> for detection of 80 known plasmid replicon marker sequences. All sequence read sets were screened against the database using SRST2 to detect the presence of these replicons in each genome. The results were transformed into a binary table in R to indicate presence/absence.

Visualization of AMR and plasmid genotypes against a core gene tree

A subtree representing the relationships between the 185 GEMS aEPEC isolates was extracted from the full core phylogeny we published previously⁶ by pruning all other tips using R packages APE⁶⁵ and GEIGER⁶⁶. The presence of acquired AMR genes, mutations and plasmid replicons was plotted as a heatmap against the phylogeny using the plotTree function for R (<https://github.com/katholt/plotTree>).

Investigation of mechanisms of AMR gene mobilization

Common AMR-associated genes that were shown to be co-occurring, specifically *bla*_{TEM-198}, *sul1*, *sul2*, *strA*, *strB*, multiple *dfxA* alleles, *tet(A)* and *tet(B)*, were further investigated in the aEPEC genome assemblies to determine if they were carried on the same mobile elements. The aEPEC genome assemblies generated previously using Velvet⁵¹ were interrogated with BLAST, using as queries the AMR genes and the sequences of the plasmids pCERC1 (accession JN012467) and pCERC2 (accession KX291024), and the transposons Tn6029 (accession CG150541)⁴⁵, Tn1721 (accession X61367)⁶⁷ and Tn10 (accession AF223162)⁶⁸. For example, if the pCERC2 backbone and AMR genes were all detected on a single contig in the genome assembly, we inferred that these genes were moving together on a pCERC2-related plasmid. Two representative aEPEC isolates that were identified as harboring a pCERC2-like plasmid backbone with different *dfx* gene insertions (strains 402635 and 400879) were selected as representatives for further analysis. These genomes were re-assembled with Unicycler⁶⁹ and annotated using Prokka⁷⁰, and compared to the reference sequences for pCERC1 and pCERC2 using BLAST. The comparison explored using Artemis Comparison Tool⁷¹ and plotted with genoplots⁷².

aEPEC isolates were inferred to be likely carriers of Tn6029 or related transposons if the entire region of *repC* to *strB* was detected by BLAST in a single contig and *bla*_{TEM-198} was also detected in the genome. (Note it is not possible for the complete transposon sequence to be assembled from short reads, since *bla*_{TEM-198} is separated from the rest of the transposon by repeat copies of IS26 which cause a break in the assembly graph). A representative strain matching this pattern (401596) was re-assembled with Unicycler⁶⁹, and the connectivity of Tn6029 genes in the resulting assembly graphs were visualized using Bandage⁷³.

The distributions of class I integrons with different cassette regions were explored by extracting the DNA sequences spanning from *int1* to *sul1* genes using BLAST and MUMmer⁷⁴. Different *dfxA* alleles were identified within the resulting sequences via BLAST searches of the ARG-ANNOT database⁵⁵. Representatives of each distinct class I integron sequence (defined by the *dfxA* gene carried) were re-assembled with Unicycler⁶⁹ and submitted to the Repository of Antibiotic-resistance Cassettes (RAC) website⁷⁵ for detailed annotation.

Antimicrobial usage data

Data on the use of antimicrobials at each of the seven GEMS sites were collected as part of the original GEMS protocol^{1,29}. These data included details of the number and type of antimicrobials prescribed to all cases presenting with watery diarrhea or dysentery at the study clinics. Two of the recorded drugs were excluded from the current analysis: pivmecillinam, because it was not used at any discernible level, and metronidazole, which is active against obligate anaerobic parasites and bacteria only⁷⁶ and therefore is not relevant to *E. coli*, which is intrinsically resistant to this agent. The frequency of prescription of each drug at each site was visualized as a heatmap in R, using the pheatmap package.

The relationship between frequencies of ciprofloxacin and trimethoprim usage and associated genetic determinants was investigated via linear regression modelling in R. For

ciprofloxacin, the genetic determinants were either one or more quinolone resistance-associated point mutations in *gyrA* (point mutations in *parC* only occurred when *gyrA* mutations were also present), or the presence of the plasmid-borne genes *qepA* or *qnrS*. Genetic determinants of trimethoprim/sulfamethoxazole resistance were the combination of at least one *dhfrA* gene together with *sul1* or *sul2*. The data were visualized in R using the *ggplot2* package³⁰.

References

1. 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. *The Lancet* **382**, 209–222 (2013).
2. Robins-Browne, R. M. *et al.* Are *Escherichia coli* pathotypes still relevant in the era of whole-genome sequencing? *Front Cell Infect Microb.* **6**, 5222–9 (2016).
3. Hernandez, R. T., Elias, W. P., Vieira, M. A. M. & Gomes, T. A. T. An overview of atypical enteropathogenic *Escherichia coli*. *FEMS Microbiol Lett* **297**, 137–149 (2009).
4. Ochoa, T. J. & Contreras, C. A. Enteropathogenic *Escherichia coli* infection in children. *Curr Opin Infect Dis* **24**, 478–483 (2011).
5. Trabulsi, L. R., Keller, R. & Gomes, T. A. T. Typical and atypical enteropathogenic *Escherichia coli*. *Emerg. Infect. Dis.* **8**, 1–6 (2002).
6. Ingle, D. J. *et al.* Evolution of atypical enteropathogenic *E. coli* by repeated acquisition of LEE pathogenicity island variants. *Nat Micro* **1**, 1–9 (2016).
7. Afset, J. E., Bevanger, L., Romundstad, P. & Bergh, K. Association of atypical enteropathogenic *Escherichia coli* (EPEC) with prolonged diarrhoea. *J Med Microbiol* **53**, 1137–1144 (2004).
8. Nguyen, R. N., Taylor, L. S., Tauschek, M. & Robins-Browne, R. M. Atypical enteropathogenic *Escherichia coli* infection and prolonged diarrhea in children. *Emerg. Infect. Dis.* **12**, 597–603 (2006).
9. Vieira, M. A. *et al.* Atypical enteropathogenic *Escherichia coli* as aetiological agents of sporadic and outbreak-associated diarrhoea in Brazil. *J Med Microbiol* **65**, 998–1006 (2016).
10. Davis, M. A. *et al.* Genotypic-phenotypic discrepancies between antibiotic resistance characteristics of *Escherichia coli* isolates from calves in management settings with high and low antibiotic use. *Appl. Environ. Microbiol.* **77**, 3293–3299 (2011).
11. Glenn, L. M. *et al.* Analysis of antimicrobial resistance genes detected in multiple-drug-resistant *Escherichia coli* isolates from broiler chicken carcasses. *Microb Drug Resist.* **18**, 453–463 (2012).
12. Medina, A. *et al.* Phenotypic and genotypic characterization of antimicrobial resistance in enterohemorrhagic *Escherichia coli* and atypical enteropathogenic *E. coli* strains from ruminants. *J Vet Diagn Invest.* **23**, 91–95 (2011).
13. Ramírez Castillo, FY. *et al.* Presence of multi-drug resistant pathogenic *Escherichia coli* in the San Pedro River located in the State of Aguascalientes, Mexico. *Front Microbiol* **4**, 1–16 (2013).
14. Scaletsky, I. C., Souza, T. B., Aranda, K. R. & Okeke, I. N. Genetic elements associated with antimicrobial resistance in enteropathogenic *Escherichia coli* (EPEC) from Brazil. *BMC Microbiology* **10**, 25 (2010).
15. Stoesser, N. *et al.* Evolutionary history of the global emergence of the *Escherichia coli* epidemic clone ST131. *MBio* **7**, e02162–15–15 (2016).
16. Magiorakos, A. P. *et al.* Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin Microbiol Infect* **18**, 268–281 (2014).
17. Lim, C. *et al.* Epidemiology and burden of multidrug-resistant bacterial infection in a developing country. *eLIFE* **5**, 1–18 (2016).
18. Musicha, P. *et al.* Trends in antimicrobial resistance in bloodstream infection isolates at a large urban hospital in Malawi (1998–2016): a surveillance study. *Lancet Infect Dis* **17**, 1042–1052 (2017).
19. Mathai, E. *et al.* Antimicrobial resistance surveillance among commensal *Escherichia coli* in rural and urban areas in Southern India. *Trop Med Int Health* **13**, 41–45 (2008).
20. Avasthi, T. S. *et al.* Genome of multidrug-resistant uropathogenic *Escherichia coli* strain NA114 from India. *J Bacteriol* **193**, 4272–4273 (2011).

21. Gupta, S., Kapur, S. & Padmavathi, D. V. Comparative prevalence of antimicrobial resistance in community-acquired urinary tract infection cases from representative states of Northern and Southern India. *J Clin Diagn Res.* **8**, 9–12 (2014).
22. Shah, M. *et al.* Prevalence, seasonal variation, and antibiotic resistance pattern of enteric bacterial pathogens among hospitalized diarrheic children in suburban regions of central Kenya. *Trop Med Health* **44**, 1–8 (2016).
23. Begum, Y. A. *et al.* Resistance pattern and molecular characterization of enterotoxigenic *Escherichia coli* (ETEC) strains isolated in Bangladesh. *PLoS ONE* **11**, e0157415–11 (2016).
24. Khalil, U. *et al.* Phenotypic and genotypic characterization of enteroaggregative *Escherichia coli* isolates from pediatric population in Pakistan. *APMIS* **124**, 872–880 (2016).
25. World Health Organization. *Antibiotic resistance: Global report on surveillance.* 1–256 (2014).
26. Traa, B. S., Walker, C. L. F., Munos, M. & Black, R. E. Antibiotics for the treatment of dysentery in children. *Int J Epidemiol.* **39**, i70–i74 (2010).
27. van Schaik, W. The human gut resistome. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **370**, 20140087–20140087 (2015).
28. Wright, G. D. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat Rev Microbiol* **5**, 175–186 (2007).
29. 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**, S232–S245 (2012).
30. Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: a new method for the analysis of genetically. *BMC Genet.* **11**, 1–15 (2010).
31. Sangare, S. A. *et al.* Very high prevalence of extended-spectrum beta-lactamase-producing Enterobacteriaceae in bacteriemic patients hospitalized in teaching hospitals in Bamako, Mali. *PLoS ONE* **12**, e0172652–11 (2017).
32. Anantham, S. & Hall, R. M. pCERC1, a small, globally disseminated plasmid carrying the *dfraA14* cassette in the *strA* gene of the *sul2-strA-strB* gene cluster. *Microb Drug Resist.* **18**, 364–371 (2012).
33. Holt, K. E. *et al.* Multidrug-resistant *Salmonella enterica* serovar paratyphi A harbors IncHI1 plasmids similar to those found in serovar typhi. *J Bacteriol* **189**, 4257–4264 (2007).
34. Chen, C.-M. *et al.* High diversity of antimicrobial resistance genes, Class 1 Integrons, and genotypes of multidrug-resistant *Escherichia coli* in beef carcasses. *Microb Drug Resist.* **23**, 915–924 (2017).
35. Oliveira-Pinto, C. *et al.* Occurrence and characterization of class 1 integrons in *Escherichia coli* from healthy individuals and those with urinary infection. *J Med Microbiol* **66**, 577–583 (2017).
36. Wong, V. K. *et al.* Phylogeographical analysis of the dominant multidrug-resistant H58 clade of *Salmonella* Typhi identifies inter- and intracontinental transmission events. *Nat Genet* **47**, 632–639 (2015).
37. Stoesser, N. *et al.* Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* isolates using whole genomic sequence data. *J. Antimicrob Chemother* **68**, 2234–2244 (2013).
38. Laxminarayan, R. *et al.* The Lancet Infectious Diseases Commission Antibiotic resistance—the need for global solutions. *Lancet Infect Dis* **13**, 1057–1098 (2013).
39. Partridge, S. R. & Hall, R. M. Evolution of transposons containing *bla*_{TEM} genes. *Antimicrob Agents Chemother* **49**, 1267–1268 (2005).
40. Bennett, P. M. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *Br J Pharmacol.* **153**, S347–S357 (2008).

41. Pallecchi, L. *et al.* Population structure and resistance genes in antibiotic-resistant bacteria from a remote community with minimal antibiotic exposure. *Antimicrob Agents Chemother* **51**, 1179–1184 (2007).
42. Andersson, D. I. & Hughes, D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat Rev Microbiol* **8**, 260–271 (2010).
43. Carattoli, A. Plasmids and the spread of resistance. *Int J Med Microbiol.* **303**, 298–304 (2013).
44. Touchon, M. *et al.* Antibiotic resistance plasmids spread among natural isolates of *Escherichia coli* in spite of CRISPR elements. *Microbiology* **158**, 2997–3004 (2012).
45. Reid, C. J., Chowdhury, P. R. & Djordjevic, S. P. Tn6026 and Tn6029 are found in complex resistance regions mobilised by diverse plasmids and chromosomal islands in multiple antibiotic resistant Enterobacteriaceae. *Plasmid* **80**, 127–137 (2015).
46. Cain, A. K., Liu, X., Djordjevic, S. P. & Hall, R. M. Transposons related to Tn1696 in IncHI2 plasmids in multiply antibiotic resistant *Salmonella enterica* serovar Typhimurium from Australian animals. *Microb Drug Resist.* **16**, 197–202 (2010).
47. Panchalingam, S. *et al.* Diagnostic microbiologic methods in the GEMS-1 case/control Study. *Clin Infect Dis.* **55**, S294–S302 (2012).
48. Morgan, D. J., Okeke, I. N., Laxminarayan, R., Perencevich, E. N. & Weisenberg, S. Non-prescription antimicrobial use worldwide: a systematic review. *Lancet Infect Dis* **11**, 692–701 (2011).
49. Sheppard, A. E. *et al.* Nested Russian doll-like genetic mobility drives rapid dissemination of the carbapenem resistance gene *bla_{KPC}*. *Antimicrob Agents Chemother* **60**, 3767–3778 (2016).
50. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom* 1–7 (2017). doi:10.1099/mgen.0.000132
51. Zerbino, D. R. & Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**, 821–829 (2008).
52. Kahlmeter, G. European harmonization of MIC breakpoints for antimicrobial susceptibility testing of bacteria. *J. Antimicrob Chemother* **52**, 145–148 (2003).
53. Sjolund-Karlsson, M. *et al.* Antimicrobial susceptibility to azithromycin among *Salmonella enterica* isolates from the United States. *Antimicrob Agents Chemother* **55**, 3985–3989 (2011).
54. Sunde, M. & Norström, M. The genetic background for streptomycin resistance in *Escherichia coli* influences the distribution of MICs. *J. Antimicrob Chemother* **56**, 87–90 (2005).
55. Gupta, S. K. *et al.* ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother* **58**, 212–220 (2013).
56. Inouye, M. *et al.* SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* **6**, 1–16 (2014).
57. Eaves, D. J. *et al.* Prevalence of mutations within the quinolone resistance-determining region of *gyrA*, *gyrB*, *parC*, and *parE* and association with antibiotic resistance in quinolone-resistant *Salmonella enterica*. *Antimicrob Agents Chemother* **48**, 4012–4015 (2004).
58. Whiteway, J. *et al.* Oxygen-insensitive Nitroreductases: analysis of the roles of *nfsA* and *nfsB* in development of resistance to 5-Nitrofurantoin derivatives in *Escherichia coli*. *J Bacteriol* **180**, 5529–5539 (1998).
59. Xavier, B. B. *et al.* Complete genome sequences of nitrofurantoin-sensitive and -resistant *Escherichia coli* ST540 and ST2747 strains. *Genome Announc.* **2**, e00239–14 (2014).
60. Jorgensen, J. H. & Ferraro, M. J. Antimicrobial susceptibility testing: a review of

- general principles and contemporary practices. *Clin Infect Dis.* **49**, 1749–1755 (2009).
61. Köser, C. U. *et al.* Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog* **8**, e1002824–9 (2012).
62. Stevenson, M. *et al.* *epiR: Tools for the analysis of epidemiological data.* (2016).
63. Csardi, G. & Nepusz, T. The igraph software package for complex network research *InterJournal Complex Systems*, 1695 (2006).
64. Carattoli, A. *et al.* *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* **58**, 3895–3903 (2014).
65. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
66. Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E. & Challenger, W. GEIGER: investigating evolutionary radiations. *Bioinformatics* **24**, 129–131 (2007).
67. Allmeier, H., Cresnar, B., Greck, M. & Schmitt, R. Complete nucleotide sequence of Tn1721: gene organization and a novel gene product with features of a chemotaxis protein. *Gene* **111**, 11–20 (1992).
68. Lawley, T. D., Burland, V. & Taylor, D. E. Analysis of the complete nucleotide sequence of the tetracycline-resistance transposon Tn10. *Plasmid* **43**, 235–239 (2000).
69. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* **13**, e1005595–22 (2017).
70. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
71. Carver, T. *et al.* Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* **24**, 2672–2676 (2008).
72. Guy, L., Kultima, J. R. & Andersson, S. G. E. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26**, 2334–2335 (2010).
73. Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**, 3550–3552 (2015).
74. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5**, 1–9 (2004).
75. Partridge, S. R., Tsafnat, G., Coiera, E. & Iredell, J. R. Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS Microbiol Rev* **33**, 757–784 (2009).
76. Samuelson, J. Why metronidazole is active against both bacteria and parasites. *Antimicrob Agents Chemother* **43**, 1533–1541 (1999).

Figures

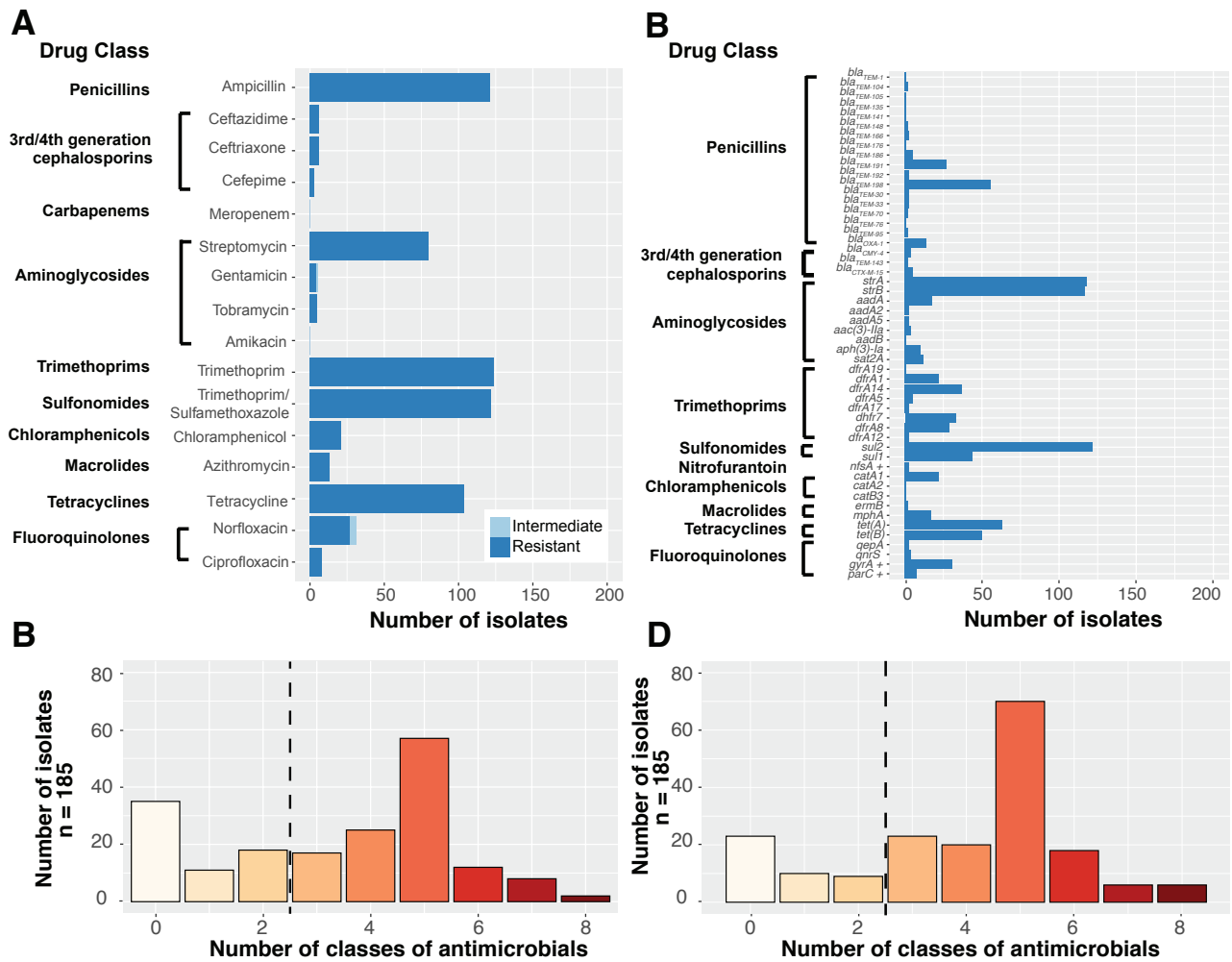


Fig 1. Prevalence of AMR-associated gene content and AMR phenotypes in 185 aEPEC isolates. Panels (A) and (B) summarize the AMR phenotypes of the aEPEC isolates. (A) AMR profiles are grouped by drug class. (B) Histogram of number of drug classes to which aEPEC strains were phenotypically resistant. Panels (C) and (D) summarize AMR-associated gene content in aEPEC strains. (C) Genes detected in the genomes associated with antimicrobial resistance are shown to the left of the graphs and are grouped by drug class. + indicates genes which have point mutations resulting in antimicrobial resistance and are not acquired through horizontal gene transfer. (D) Histogram of number of classes to which aEPEC strains were detected as having AMR-associated gene content.

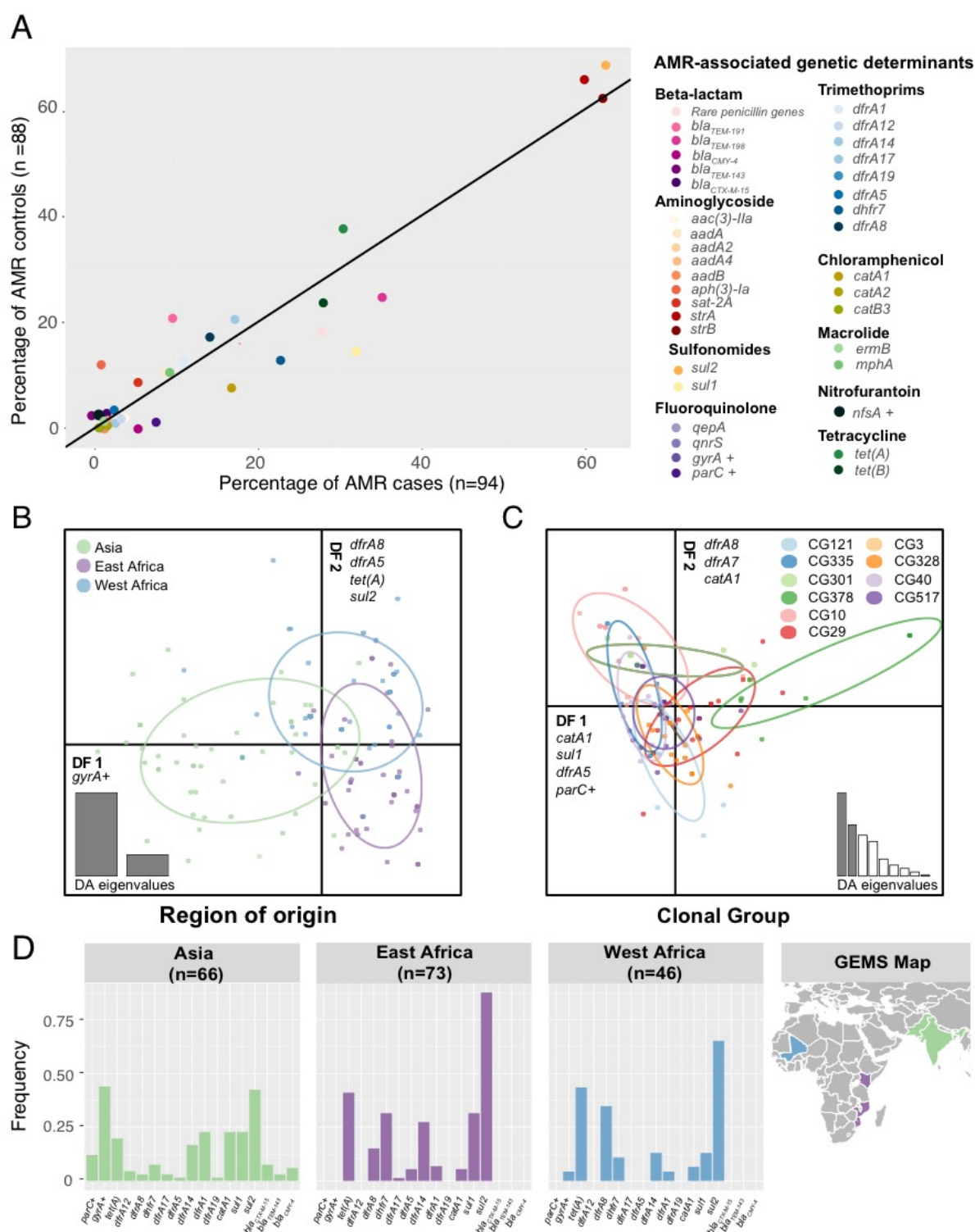


Fig 2. AMR gene content is explained by region of isolation not disease status or clonal group.

(A) Frequencies of AMR-associated genes in aEPEC by case or control status. Genes encoding AMR are shown to the right of the graph and are grouped by drug class. (B–C) Discriminant analysis of principal components (DAPC) based on known genetic determinants of AMR. The graphs display the discriminant functions (DF) that best discriminate isolates into region of isolation (B) or clonal group (C). Data points are colored based on their demographic group (according to inset legend) and the genetic determinants most correlated with the DFs are labeled on the DF axes. (D) Frequency of AMR genetic determinants that differed between Asia, East Africa and West Africa. + indicates these genes have point mutations resulting in AMR and are not acquired through horizontal gene transfer.

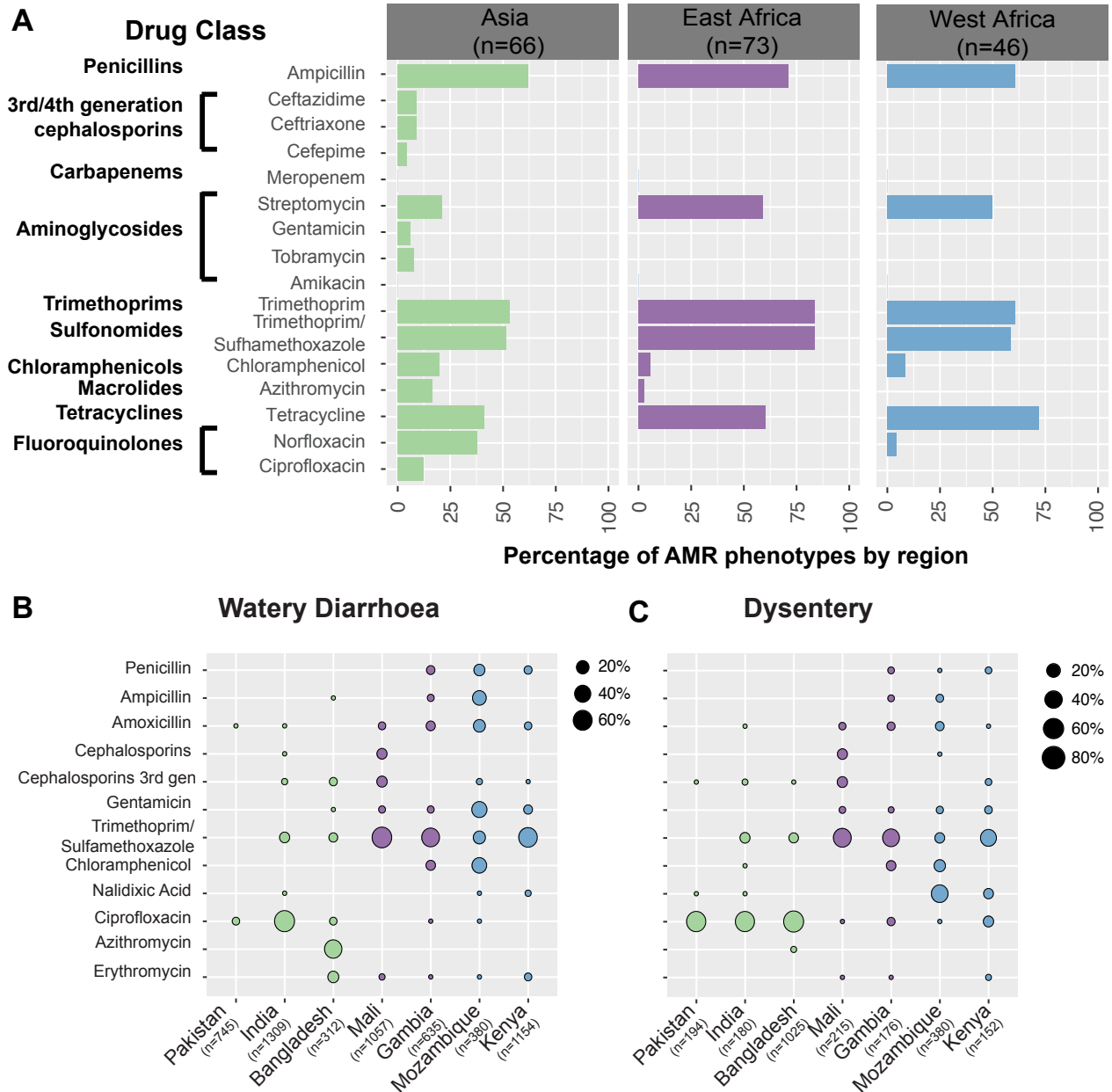


Fig 3. AMR phenotypes by region and antimicrobial use at study sites. (B) AMR phenotypes of GEMS isolates stratified by region of isolation. (B-C) Percentage of antimicrobials prescribed to patients with watery diarrhea (B) or dysentery (C) at each of the seven sites.

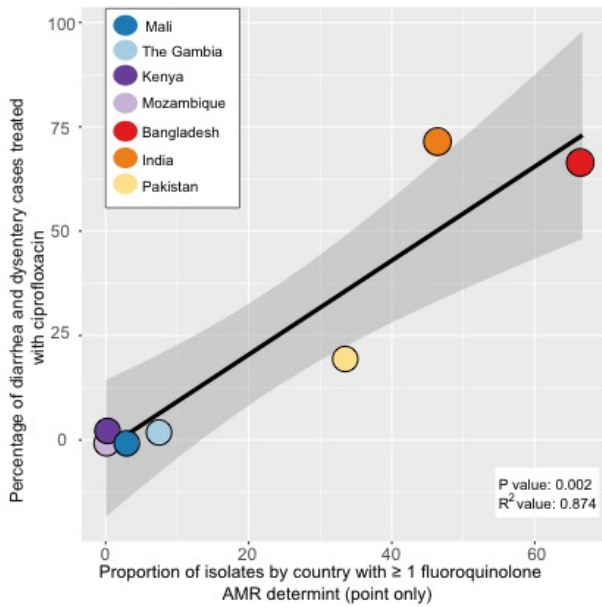


Fig 4. Relationship between use of ciprofloxacin at GEMS study sites and ciprofloxacin resistance in aEPEC. Linear regression (mean \pm 95% CI) of proportion of aEPEC with genetic determinants of ciprofloxacin resistance (from point mutations only) vs the proportion of watery diarrhea cases and dysentery case treated with ciprofloxacin at GEMS study sites.