Modulation of multidrug-resistant clone success in Escherichia coli populations

Pöntinen, Anna K; Gladstone, Rebecca A; Pesonen, Henri; Pesonen, Maiju; Cléon, François; Parcell, Benjamin J; Kallonen, Teemu; Simonsen, Gunnar Skov; Croucher, Nicholas J; McNally, Alan; Parkhill, Julian; Johnsen, Pål J; Samuelsen, Ørjan; Corander, Jukka

DOI:
10.1016/S2666-5247(23)00292-6

Modulation of multidrug-resistant clone success in *Escherichia coli* populations: a longitudinal, multi-country, genomic and antibiotic usage cohort study

Anna K Pöntinen, Rebecca A Gladstone*, Henri Pesonen*, Maiju Pesonen, François Cléon, Benjamin J Parcell, Teemu Kallonen, Gunnar Skov Simonsen, Nicholas J Croucher, Alan McNally, Julian Parkhill, Pål J Johnsen, Ørjan Samuelsen, Jukka Corander*

Summary

Background The effect of antibiotic usage on the success of multidrug-resistant (MDR) clones in a population remains unclear. With this genomics-based molecular epidemiology study, we aimed to investigate the contribution of antibiotic use to *Escherichia coli* clone success, relative to intra-strain competition for colonisation and infection.

Methods We sequenced all the available *E coli* bloodstream infection isolates provided by the British Society for Antimicrobial Chemotherapy (BSAC) from 2012 to 2017 (n=718) and combined these with published data from the UK (2001–11; n=1090) and Norway (2002–17; n=3254). Defined daily dose (DDD) data from the European Centre for Disease Prevention and Control (retrieved on Sept 21, 2021) for major antibiotic classes (β-lactam, tetracycline, macrolide, sulfonamide, quinolone, and non-penicillin β-lactam) were used together with sequence typing, resistance profiling, regression analysis, and non-neutral Wright–Fisher simulation-based modelling to enable systematic comparison of resistance levels, clone success, and antibiotic usage between the UK and Norway.

Findings Sequence type (ST)73, ST131, ST95, and ST69 accounted for 892 (49·3%) of 1808 isolates in the BSAC collection. In the UK, the proportion of ST69 increased between 2001–10 and 2011–17 (p=0·0004), whereas the proportions of ST73 and ST95 did not vary between periods. ST131 expanded quickly after its emergence in 2003 and its prevalence remained consistent throughout the study period (apart from a brief decrease in 2009–10). The extended-spectrum β-lactamase (ESBL)-carrying, globally disseminated MDR clone ST131–C2 showed overall greater success in the UK (154 [56·8%] of 271 isolates in 2003–17) compared with Norway (51 [18·3%] of 278 isolates in 2002–17; p<0·0001). DDD data indicated higher total use of antimicrobials in the UK, driven mainly by the class of non-penicillin β-lactams, which were used between 2·7-times and 5·1-times more in the UK per annum (ratio mean 3·7 [SD 0·8]). This difference was associated with the higher success of the MDR clone ST131–C2 (pseudo-R^2 69·1%). A non-neutral Wright–Fisher model replicated the observed expansion of non-MDR and MDR sequence types under higher DDD regimes.

Interpretation Our study indicates that resistance profiles of contemporaneously successful clones can vary substantially, warranting caution in the interpretation of correlations between aggregate measures of resistance and antibiotic usage. Our study further suggests that in countries with low-to-moderate use of antibiotics, such as the UK and Norway, the extent of non-penicillin β-lactam use modulates rather than determines the success of widely disseminated MDR ESBL-carrying *E coli* clones. Detailed understanding of underlying causal drivers of success is important for improved control of resistant pathogens.

Funding Trond Mohn Foundation, Marie Skłodowska–Curie Actions, European Research Council, Royal Society, and Wellcome Trust.

Introduction

Escherichia coli is a globally leading opportunistic pathogen that ubiquitously colonises the gut of humans and other mammals.[1] Bloodstream infections caused by extra-intestinal pathogenic *E coli* are of serious clinical concern,[2] particularly due to a notable increase in both incidence and the fraction of isolates harbouring a multidrug-resistant (MDR) phenotype among all infections related to extra-intestinal pathogenic *E coli* over the past two decades. In a seminal study, Goossens and colleagues demonstrated a significant positive association between the outpatient use of antibiotics and the extent of antimicrobial resistance (AMR) among *E coli* infections on a Europe-wide scale.[3] The selective effect of antibiotics on gut microbes can hardly be debated. However, the importance of antibiotic use in relation to other factors determining the success of *E coli* clones in a population of hosts remains unclear.

Several large-scale longitudinal genomic cohort studies of *E coli* from bloodstream infection in Europe have sought to establish the population frequencies of both clones and

Lancet Microbe 2024

Published Online https://doi.org/10.1016/S2666-5247(23)00292-6

*Contributed equally

Department of Biostatistics, Faculty of Medicine, University of Oslo, Oslo, Norway (A K Pöntinen PhD, R A Gladstone PhD, H Pesonen PhD, M Pesonen PhD, Prof J Corander PhD); **Norwegian National Advisory Unit on Detection of Antimicrobial Resistance, Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway** (A K Pöntinen, Prof Ø Samuelsen PhD); **Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital Research Support Services, Oslo, Norway** (M Pesonen); **Department of Pharmacy** (F Cléon PhD, Prof P J Johnsen PhD, Prof Ø Samuelsen) **and Research Group for Host–Microbe Interaction** (Prof G S Simonsen PhD), **Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway; Medical Microbiology, Ninewells Hospital and Medical School, Dundee, UK** (B J Parcell FRCPath); **Department of Clinical Microbiology, Turku University Hospital, Turku, Finland** (T Kallonen PhD); **Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway** (Prof G S Simonsen); **MRC Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, Imperial College London, London, UK** (N J Croucher PhD); **Institute of Microbiology and Infection, University of Birmingham, Birmingham, UK** (Prof A McNally PhD); **Department of Veterinary**

Medicine, University of Cambridge, Cambridge, UK (Prof J Parkhill PhD)**; Parasites and Microbes, Wellcome Sanger Institute, Cambridge, UK** (Prof J Corander)**; Helsinki Institute of Information Technology, Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland** (Prof J Corander)

Correspondence to:
Dr Anna K Pöntinen, Department of Biostatistics, Faculty of Medicine, University of Oslo, Oslo 0317, Norway
a.k.pontinen@medisin.uio.no

or

Prof Jukka Corander, Department of Biostatistics, Faculty of Medicine, University of Oslo, Oslo 0317, Norway
jukka.corander@medisin.uio.no

Research in context

Evidence before this study

Antimicrobial resistance surveillance data have established a significant positive association between usage of antibiotics and the population frequency of resistance in major bacterial pathogens across different countries. However, for opportunistic pathogens normally residing in the gut, such as *Escherichia coli*, the picture remains more complicated and there has been debate about the relative contributions of different drivers of resistance. Different resistance mechanisms can, in general, show strikingly different trends even within a single population and it is necessary to consider separately the roles of de novo resistance mutations arising under treatment and selection of lineages already carrying resistance genes, such as extended-spectrum β-lactamases (ESBLs), leading to clonal expansions. We searched PubMed from database inception to Oct 9, 2022, with the search terms "coli" AND "ST131" AND "infection" AND "bloodstream" AND "whole-genome" AND "sequencing" AND "longitudinal". Searches were restricted to primary research articles published in English. This search returned four publications, of which the relevant ones for the current study were the Norwegian genomic surveillance study covering 2002–17 and the UK genomic surveillance study covering 2002–11. Longitudinal *E coli* genome data from these studies have shown both a stable coexistence of susceptible and resistant lineages for many different classes of antibiotics, and clonal expansions of less resistant lineages, indicating that factors not related to resistance are important for the success of these bacteria. Negative frequency-dependent selection (NFDS) of clone-specific gene content has been suggested as a mechanism for maintaining the stable coexistence of *E coli* clones. It remains unclear to what extent the clinical use of antibiotics alters population trajectories of clones under the NFDS hypothesis.

Added value of this study

To our knowledge, our study presents the first possibility to directly compare the population trends of *E coli* between two countries (Norway and the UK) and to explain differences as a function of country-wise antibiotic usage levels. This is achieved by generating unbiased population genomic data to establish two synchronous longitudinal *E coli* bloodstream infection cohorts covering almost two decades. The data further enable separation of stationary, transient, and expanding trends of resistance levels in the context of antibiotic usage.

Implications of all the available evidence

Combining the NFDS hypothesis with selection pressure from antibiotic use suggests that higher usage of specific β-lactam antibiotics, such as cephalosporins, drives the equilibrium population frequency upwards for successful clones with a stable carriage of ESBLs, such as the sequence type (ST)131–C2 clade. Fluoroquinolone use and resistance was similar between the two countries by 2017, but the UK had a significantly higher proportion of ciprofloxacin resistant strains, belonging to ST131–C2. The level of inferred trimethoprim resistance was also higher in this clade in the UK than in Norway, and both of these differences are explained by the resistance elements common in ST131–C2. However, there was no systematic difference in resistance between Norway and the UK across the three other top-ranking clones of *E coli*, despite smaller trimethoprim defined daily dose levels in Norway after 2005. The data thus indicate that benefits gained by bacteria from antimicrobial resistance do not manifest themselves uniformly across different clones or classes of antibiotics.

AMR for the clinically relevant antibiotic classes.[4–7] A consensus finding from these studies is that the same dominant clones are found circulating over time in the populations and that their resistance profiles are largely similar across countries (France, Norway, and the UK). Among the dominant clones, sequence type (ST)73 and ST95 are examples of circulating *E coli* with less resistance, few MDR isolates, and the near absence of extended-spectrum β-lactamase (ESBL) genes. By contrast, the clones ST69 and ST131, which quickly increased in frequency in the early 2000s and have been referred to as pandemic clones, both contribute to the burden of MDR infections. In particular, the ST131 clone is the most widely disseminated MDR extra-intestinal pathogenic *E coli* clone globally. Of note, the ST131 lineage consists of four identified major clades—A, B, C1, and C2—each of which has been found circulating widely. Earlier studies have established that the C2 clade is predominantly ESBL positive, whereas the other three clades show a variable degree of the MDR phenotype, ranging from zero observed prevalence in B to variable levels in C1.[4,5,8] A Norwegian longitudinal study (Norwegian Surveillance Programme on Resistant Microbes, NORM) identified numerous independent de novo acquisitions of ESBL genes within the A clade after 2010.[4] However, it is unclear how much these might have contributed to the overall success of the A clade, which by then had already reached a fairly stable population frequency in Norway. An earlier longitudinal study in the UK covering the years 2001 to 2011 (British Society for Antimicrobial Chemotherapy, BSAC) found similar expansions to the Norwegian study.[5] However, a key observed difference was the reversed ranking of population frequencies of A and C2 clades between the two countries.[4]

Longitudinal *E coli* genome data sampled in an unbiased manner have shown both a stable coexistence of predominantly susceptible and resistant lineages for many different classes of antibiotics, and clonal expansions of lineages with strikingly distinct MDR prevalence, indicating that factors not related to resistance are also important for the success of these bacteria. Multi-locus negative frequency-dependent selection (NFDS) has been suggested as a mechanism for maintaining stable population frequencies of *E coli* clones.[9] Both the BSAC and NORM studies empirically demonstrated a pattern of transient disruption and rapid return to

equilibrium population frequencies, which bears the hallmarks of NFDS in general, irrespective of what the underlying mechanism of balancing selection was. However, it remains unclear to what extent clinical use of antibiotics could alter the population trajectories of clones under the NFDS hypothesis, assuming that over a relatively short timescale, most selective benefits would stem from an increased competitiveness of isolates with a pre-existing resistance phenotype.

To answer these open questions, we extended the previous BSAC collection (2001–11) from Kallonen and colleagues[5] (referred to as BSAC1 in the current study) by generating genome sequences from bloodstream infection isolates from the BSAC collection covering the years 2012 to 2017 (BSAC2) and comparing these to the published NORM collection (2002–17). The two parallel unbiased collections covering nearly two decades enabled us to perform, for the first time, a cross-country analysis of the clonal diversity of *E coli* from bloodstream infection and to investigate the roles of both antibiotic use and balancing selection acting on the dominant clones in these two populations. In particular, we sought to assess how defined daily doses (DDDs) for relevant major classes of antibiotics are associated with the prevalence of clones and their AMR frequencies.

Methods

Study design

The analyses included a total of 1808 *E coli* isolates associated with bacteraemia from the BSAC Bacteraemia Resistance Surveillance Programme in the UK. We retrieved whole-genome sequences of 1090 isolates collected in 2001–11 from Kallonen and colleagues (BSAC1).[5] In addition, within this study, we sequenced the genomes of 718 isolates collected in 2012–17 (BSAC2; appendix 1 p 1, appendix 2 pp 1–2). For comparative purposes, whole-genome sequences of 3254 *E coli* bloodstream infection isolates from the NORM study,[4] collected in 2002–17, were included in the analyses (appendix 2 pp 1–2). Ethics approval was not needed because the study included no human samples nor any sensitive data. Data from the BSAC and NORM cohort studies are de-identified.

Procedures

Isolates from BSAC2 were sequenced on the Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, USA) with 384-plexing and 150-bp paired-end reads (appendix 2 p 2). All sequence data were assembled and annotated with the same published pipeline with default parameters.[10–12] Multi-locus sequence types were retrieved using SRST2 v0.2.0, and the collections were subsequently characterised by multi-locus sequence type over time; changes in the proportions of the four major sequence types (ST69, ST73, ST95, and ST131) between 2001–10 and 2011–17 in the UK were estimated and compared with the data from Norway.[4] ST131 clades were designated according to *fimH* types and clade-specific single nucleotide polymorphisms,[8,13,14] and the clade composition was characterised in the UK and Norway and between time periods (2001–10 and 2011–17). Invasiveness in contrast to carriage of the ST131 clades was evaluated by risk ratios (RRs). AMR data were retrieved by screening genes and point mutations from genome assemblies using AMRFinderPlus v3.10.18 with database v2022-05-26.1 and *Escherichia* taxonomy group (appendix 2 p 2) and by phenotypic antimicrobial susceptibility testing (appendix 1 pp 2–8): $bla_{\text{CTX-M}}$ genes were used as a proxy for ESBL production and third-generation cephalosporin resistance,[4] and *dfr* genes were considered as representative for trimethoprim resistance.[15] Ciprofloxacin resistance designation was based on antimicrobial susceptibility testing using European Committee on Antimicrobial Susceptibility Testing (EUCAST) v13 2023 breakpoints. Overall $bla_{\text{CTX-M}}$ prevalence, including the different variants and by ST131 clade, was measured between time periods (2003–10 and 2011–17).

For more on **EUCAST** see https://www.eucast.org

To compare antimicrobial usage in the UK and Norway and between time periods (1997–2007 and 2008–17) within the countries, usage data for major antibiotic classes (β-lactam, tetracycline, macrolide, sulfonamide, quinolone, and non-penicillin β-lactam) were retrieved from the European Centre for Disease Prevention and Control database (retrieved on Sept 21, 2021; appendix 2 pp 2–3, 12–13).[16] Rate of consumption is expressed in DDD per 1000 individuals, and antimicrobial classes categorised according to the anatomical therapeutic chemical (ATC)/DDD Index. The data included antibacterials for systemic use (ATC group J01) in the community (primary care sector).

Multi-locus NFDS has been suggested as a mechanism for maintaining stable population frequencies of *E coli* clones by balancing selection that acts on accessory loci and constrains their population prevalence over time through selective disadvantage with an increasing prevalence over the equilibrium.[9] Therefore, a non-neutral Wright–Fisher NFDS simulation-based modelling framework was employed to assess the interplay of antibiotic use and intra-strain competition in colonisation (figure 1). In this model, selection acts on standing variation present in a finite population with non-overlapping generations. Effects of de novo mutations are excluded due to the relatively short timescale considered in the model. Each individual strain is represented by a stochastic number of isolates. Each strain has an AMR phenotype profile that is modelled separately and assigned a distribution that determines the relative frequency of resistance gene-carrying isolates for the strain. After initialising the population with a random state drawn from a given distribution, the population undergoes transitions from one generation, t, to the next ($t \rightarrow t+1$), during which survival (reproduction) of an isolate is determined by: the specified NFDS mechanism (appendix 2 p 3); and resistance to an antibiotic, should the isolate be subject to an antibiotic treatment in that generation. Antibiotic treatment mimicking use of cephalosporins is applied to randomly chosen isolates in the population with a constant pressure set by a chosen DDD level. Isolates carrying the $bla_{\text{CTX-M}}$ locus are

See **Online** for appendix 1

See **Online** for appendix 2

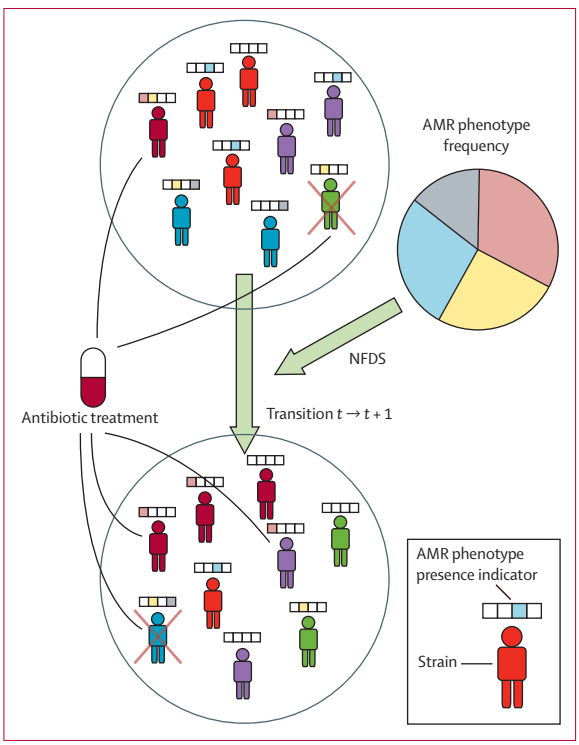

Figure 1: **Schematic of the non-neutral Wright–Fisher NFDS simulator model**
The colours of cartoon hosts represent individual strains, and each strain has an AMR phenotype profile, depicted by the bar above each cartoon host. $t \rightarrow t+1$ marks the transition from one generation to the next, during which survival (reproduction) of an isolate is determined by the specified NFDS mechanism and resistance to an antibiotic, if the isolate is subject to an antibiotic treatment in that generation. Isolates carrying the $bla_{CTX\text{-}M}$ locus (red box) are defined as resistant to the treatment. The model is described in detail in the Methods and in appendix 2 (p 3). NFDS=negative frequency-dependent selection. AMR=antimicrobial resistance.

defined as resistant to the treatment, whereas those lacking the locus will be erased by it. In addition, a proportion of isolates in the simulated population are replaced by migration from an external source to avoid stochastic extinction of strains (appendix 2 p 3).

The simulation started from all lineages residing at their assumed population-specific NFDS equilibrium, except for ST69 and ST131, which both started from a low frequency, with a fitness advantage that drove an expansion mimicking that observed in the surveillance datasets. In addition, a beta-regression analysis was used to quantify the significance and amount of variation in clone frequencies explained by DDD variation (appendix 2 pp 3–4).

Statistical analysis

RRs with 95% CIs were applied as a measure of invasiveness using the ST131 clade frequencies in carriage, as described by Mäklin and colleagues,[17] and in disease, according to the BSAC collections. To determine the significance of differences in proportions within strains and selected resistance patterns between countries (the UK *vs* Norway) and time periods (2001–10 or 2003–10 *vs* 2011–17), Fisher's exact p value was used when the expected cell count was less than five, and otherwise the χ^2 p value was reported. The differences in the mean proportions of *dfr* were assessed using a two-way ANOVA for rank-transformed proportions. The Mann-Whitney *U* test with continuity correction was used to compare the total DDD values between countries and time periods (1997–2007 *vs* 2008–17). Positive and negative predictive values of *dfr* genes as a proxy for trimethoprim resistance were calculated using standard equations, where positive predictive value equals probability of presence when the test is positive and negative predictive value equals probability of absence when the test is negative. Statistical analyses were performed in R v4.2.2 or later (appendix 2 p 4).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

The combined BSAC collection (2001–17) included 218 known multi-locus sequence types. The four largest sequence types, ST73, ST131, ST95, and ST69, accounted for almost half of the total number of isolates (892 [49·3%] of 1808 isolates; figure 2). The proportion of ST69 increased from 45 (4·5%) of 990 isolates in 2001–10 to 72 (8·8%) of 818 isolates in 2011–17 (p=0·0004), whereas ST73 and ST95 showed no significant changes within the population between these time periods. ST131 emerged in the BSAC collection in 2003 and then quickly expanded; it showed a brief rapid decline and recovery around 2009–10, similar to the data from a local study of *E coli* bloodstream infection in Oxfordshire during 2008–18.[7] In the BSAC collection, ST131–C2 clade isolates accounted for 154 (56·8%) of the total 271 ST131 isolates in 2003–17, and their proportion in the collection remained stable when compared between 2001–10 and 2011–17 (p=0·41), as did that of clade B (p=0·71; figure 3A). However, clade A increased slightly from seven (0·7%) of 990 isolates to 21 (2·6%) of 818 isolates (p=0·0027) and clade C1 increased slightly from 11 (1·1%) of 990 isolates to 27 (3·3%) of 818 isolates (p=0·0022), from 2001–10 to 2011–17. An opposite population frequency of clades A and C2 was seen in Norway, where clade A was more common and C2 less common than in the UK (figure 3B), with ST131-C2 accounting for 51 (18·3%) of the total 278 ST131 isolates in Norway in 2002–17.[4] Clades B0 and C0 remain rare and included in total only three and nine isolates, respectively, in the combined BSAC collection.

Use of antimicrobials per country and over the years 1997 to 2017 was quantified within the primary care sector as DDDs by antimicrobial class categorised according to the ATC/DDD index (figure 4A, B; appendix 2 pp 12–13). The total use of antimicrobials in the UK (mean 15·2 DDD per 1000 population [SD 2·0]) in 1997–2017 significantly exceeded usage in Norway (mean 12·7 DDD per 1000 population [0·8]; mean difference 2·5 [95% CI 1·4–3·8]; p<0·0001; appendix 2 p 14). The largest difference was

found in the use of non-penicillin β-lactams (J01D), with annual UK DDDs ranging between 2·7-times and 5·1-times those of Norway (ratio mean 3·7 [SD 0·8]; UK mean 0·6 DDD per 1000 population [SD 0·3]; Norway mean 0·2 DDD per 1000 population [0·1]). Although a slightly decreasing trend was observed from 2014 onwards, the total use of antimicrobials in the UK still increased in 2008–17 compared with 1997–2007 (p<0·0001), whereas the total DDDs in Norway exhibited a slight decrease (p=0·012) in the later years. The extent of penicillin use can be excluded as a potential driver of higher success of MDR in the UK because DDD values in Norway are slightly higher than in the UK for this class during the first 10 years of the data.

Using known *dfr* genes as a proxy for trimethoprim resistance in the NORM collection had a positive predictive value of 90·3% (95% CI 88·1–92·3) and a negative predictive value of 97·3% (95% CI 96·6–97·9). Fractions of resistant isolates varied considerably over the study years but the pattern was generally similar between the countries, with resistance being much more common in ST69 and ST131 than in ST73 and ST95 (appendix 2 p 6). For these four top-ranking clones, the rank-transformed proportions of *dfr* positive isolates averaged over 2002–17 were found to be significantly different between the UK and Norway only within ST131, with the UK having a higher proportion of *dfr*-positive isolates (mean difference on a rank scale of 22·2 [95% CI 6·6–37·8]; p=0·0057; appendix 2 p 7). Use of antibiotics in this class was on average higher in the UK, but this did not translate to significantly higher levels of resistance in ST69, ST73, or ST95.

In the combined BSAC collection, 128 (7·1%) of 1808 isolates harboured a $bla_{CTX\text{-}M}$ gene. An increase was observed from 53 (6·6%) of 809 isolates in 2003–10 to 75 (9·2%) of 818 isolates in 2011–17, although the change was not significant (p=0·062). Altogether, we detected eight different CTX-M variants, of which CTX-M-15 was the most abundant (115 [89·8%] of 128 positive isolates). Similar to Norway,[4] in the UK, ST131 was the largest contributor to CTX-M positive isolates (96 [75·0%] of 128 isolates), and of these, most (85 [88·5%] of 96 isolates) belonged to ST131–C2 (appendix 2 p 8). No CTX-M positive isolates were found in the BSAC collection in 2001–02. ST131–C2 was also the largest contributor to ciprofloxacin resistance (appendix 2 pp 5, 9).

The ST131 clades differ not only in resistance, but also in their relative frequencies in carriage and disease. Prevalence in carriage and the invasiveness of the individual strains will affect the dynamics observed in disease. We used healthy *E coli* colonisation frequencies from a published UK neonatal cohort to assess risk ratios for clone virulence in order to compare success in disease against colonisation.[17] Using the combined BSAC collections we were able to re-estimate RRs, a measure of invasiveness, by improving the temporal matching of the carriage and disease collection to provide more robust estimates of invasiveness (appendix 2 p10). Regardless of the disease collection used, clade C2 RRs significantly exceed 1 (BSAC 2003–17 RR 4·5 [95% CI 2·3–8·8]; p<0·0001), indicating that they are found considerably more often than expected in disease. Conversely, clades A and B with estimated RRs less than 1 tend to be found more often in carriage; the RR for clade A was significantly below 1, whereas for clade B the change was not significant (BSAC 2003–17 clade A RR 0·5 [0·3–0·9]; p=0·021; BSAC 2003–17 clade B RR 0·7 [0·4–1·4]; p=0·44).

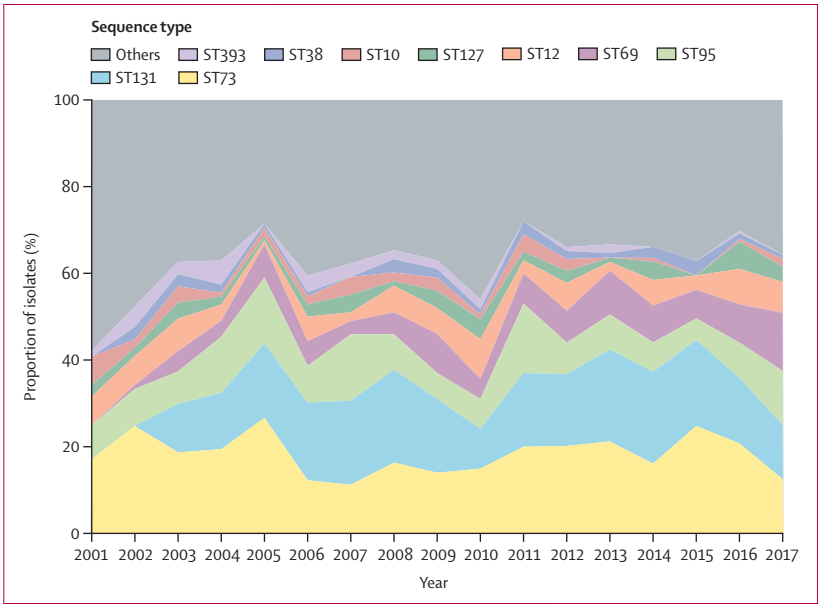

Figure 2: **Proportions of major sequence types in the combined British Society for Antimicrobial Chemotherapy collection (n=1808) in 2001–17**
The nine largest sequence types are shown with colours as indicated in the key; less frequent sequence types are combined and denoted as others, coloured in grey. ST=sequence type.

To quantify how much of the variation in the prevalence of ST131–C2 is explained by variation in DDD for the class of other β-lactams, we fitted a beta-regression model with logit-link function and DDD, country, and the interaction of DDD and country as explanatory variables (appendix 2 p 4). The model explained a substantial proportion of variation in ST131–C2 prevalence (pseudo-R^2 69·1%).

Given that ST69 and ST131 were the two lineages observed with rapid expansion in bloodstream infection cases starting around the mid-2000s, we sought to quantify the relative expansion of them under the multi-locus NFDS hypothesis and different DDD regimes. The temporal change in population frequencies for the four largest lineages (ST69, ST73, ST95, and ST131) common to the BSAC and NORM collections is shown in appendix 2 (p 11). As expected, increasing DDD level led to a monotonically increasing relative frequency of ST131 under both BSAC and NORM population models, because ESBL carriage is common in this lineage and it confers a fitness advantage in the model. No similar trend was observed for ST69, which expanded to a similar population frequency in the BSAC and NORM collections. Given the lack of ESBL genes, the ST69 expansion ability was largely determined by the NFDS component of the model.

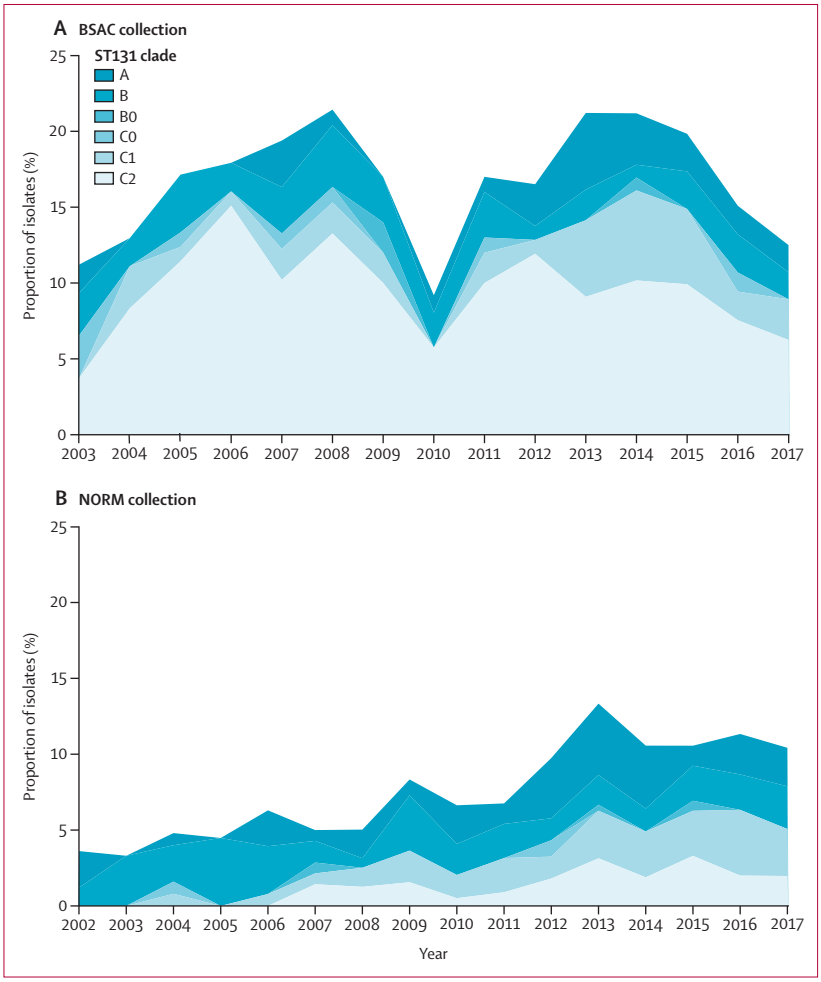

Figure 3: **Proportion of ST131 clades over time**
(A) ST131 isolates in the combined BSAC collection from the UK (n=1808) in 2003–17. No ST131 isolates were detected in 2001–02. (B) ST131 isolates in the NORM collection from Norway (n=3254) in 2002–17. Data for part B were retrieved from Gladstone and colleagues.[4] BSAC=British Society for Antimicrobial Chemotherapy. NORM=Norwegian Surveillance Programme on Resistant Microbes. ST=sequence type.

We sought to investigate in closer detail the frequency changes related to ST131 under the different modelling scenarios. The relative frequencies of ST131 clades A, B, C1, and C2 at differing levels of DDD are shown in figure 5. The findings were as expected; for example, the frequency of the B clade decreased with increasing DDD level. This clade has not been observed to carry any ESBL genes in either collection and it consequently suffered an increasing fitness disadvantage under an increasing DDD level. Because the average empirical population frequency of C2 was around three-times higher in the BSAC collection than in NORM, this clade would, by definition, always tend to a higher frequency in the BSAC simulation compared with the NORM simulation. We assumed that the stable population frequency of C2 in the UK and Norway was affected by both total use of relevant antibiotics and the intrinsic factors determining the fitness of the clade in colonisation competition. Therefore, we aimed to disentangle the effects of the NFDS and DDD selection, by letting the BSAC population evolve towards the empirical NORM equilibrium frequencies while being subject to the different DDD selection levels (BSAC scenario 2 in figure 5). In this simulation scenario, the starting population had a higher prevalence of the ESBL gene but the equilibrium frequency of the gene would tend to that of NORM under no or small antibiotic selection effect, as was seen to happen with the baseline DDD level (top row in figure 5). By contrast, under the two highest DDD levels, the equilibrium frequency of C2 clearly surpassed that of the A and C1 clades. This was a marked deviation from the equilibrium observed under the NORM scenario, where the three clades A, C1, and C2 were more uniformly distributed.

Discussion

The increasing public health burden of bloodstream infections caused by extra-intestinal pathogenic *E coli* warrants particularly high-resolution epidemiological scrutiny of the interplay of ecology and evolution in shaping the population of pathogens causing these infections. Our study aimed to achieve this through use of an unbiased selection of isolates, longitudinal surveillance, and high-quality whole-genome sequencing. By leveraging nationwide surveillance systems from two neighbouring countries with overall similar levels of antibiotic usage and clone distributions, we sought to maximise the opportunity to robustly identify reasons for clone success, while minimising the effect of confounding factors by careful study design. However, uncontrolled confounding, such as differing international travel patterns between the study countries, might also have contributed to the higher prevalence of ST131–C2 in the UK via more extensive import and constitutes the main limitation in the interpretation of our results. However, full causal attribution would only be possible in a randomised trial setting, which is not feasible in the current population-wide context.

Extension of the earlier BSAC data confirms that the differences between Norway and the UK observed by Gladstone and colleagues in 2002–11[4] remained the same during the later years 2012–17. Furthermore, our findings reinforce a previous quantification of the relative virulence of different *E coli* clones from 2022, which established a number of clones as significantly less or significantly more virulent than baseline by comparing clone frequencies in asymptomatic neonatal colonisation to those observed in bloodstream infections.[17] Our new data demonstrate that the UK-based estimates of invasiveness were robust and further highlight the importance of geographically and temporally matched collections in comparisons of colonisation and infection. An important conclusion from these estimates is that ST131 clade A is considerably more successful in colonisation of hosts than is clade C2, whereas the converse is true for bloodstream infections. This finding is well aligned with the much higher population expansion rate estimate of ST131 clade A for the years 2004–06 (around ten-times higher) combined with the same observed order of magnitude in the bloodstream infection frequency,[4]

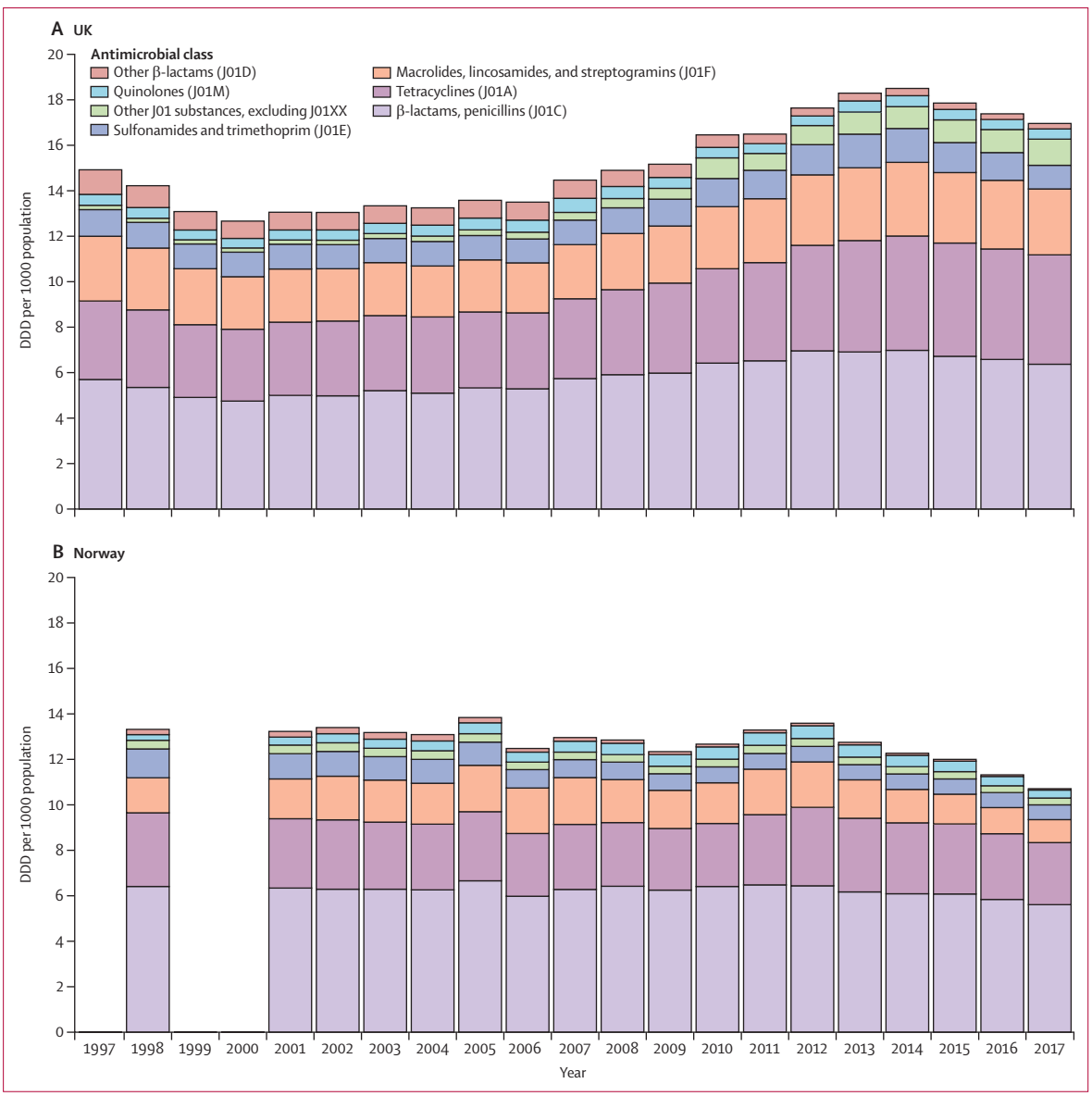

Figure 4: **Rate of antimicrobial consumption in the community (primary care) as DDD per 1000 population in the UK (A) and Norway (B), categorised according to the ATC/DDD Index, in 1997–2017**

No data were available for Norway in 1997 and 1999–2000. Group J01XX (including methenamine) is excluded from other J01 substances (green) in both the UK and Norway datasets. ATC=anatomical therapeutic chemical. DDD=defined daily dose.

suggesting that clade A did spread both faster and further among the host population with a larger effective population size, but was less likely to cause bloodstream infections. We also previously reported that clade A shared a common ancestor with the rest of ST131 in around 1977, and clade C2 probably diverged from C1 in the 1990s.[4] A relatively rapid specialisation of the clades of ST131 seems to have occurred over the past 50 years, resulting in the differences in AMR, population size, invasiveness, plasmids, and antigens (*fimH* and O-H-type) that characterise them today.

Our study lends further support to the observed importance of non-antibiotic-related balancing selection in maintaining an equilibrium among clones.[4,5,9] However, it also highlights that the selective effects of antibiotics at the population level vary considerably across different classes of broad-spectrum antibiotics, and that particular classes, such as cephalosporins belonging to non-penicillin β-lactams, exhibit the power to further modulate the success of MDR clones, such as the C2 clade of ST131. Of the top four clones, trimethoprim resistance was significantly more common only within ST131 in the UK than in Norway, which could be explained by co-selection within the C2 clade, as previously suggested in Sweden.[18] However, our data demonstrate that selective effects do not necessarily

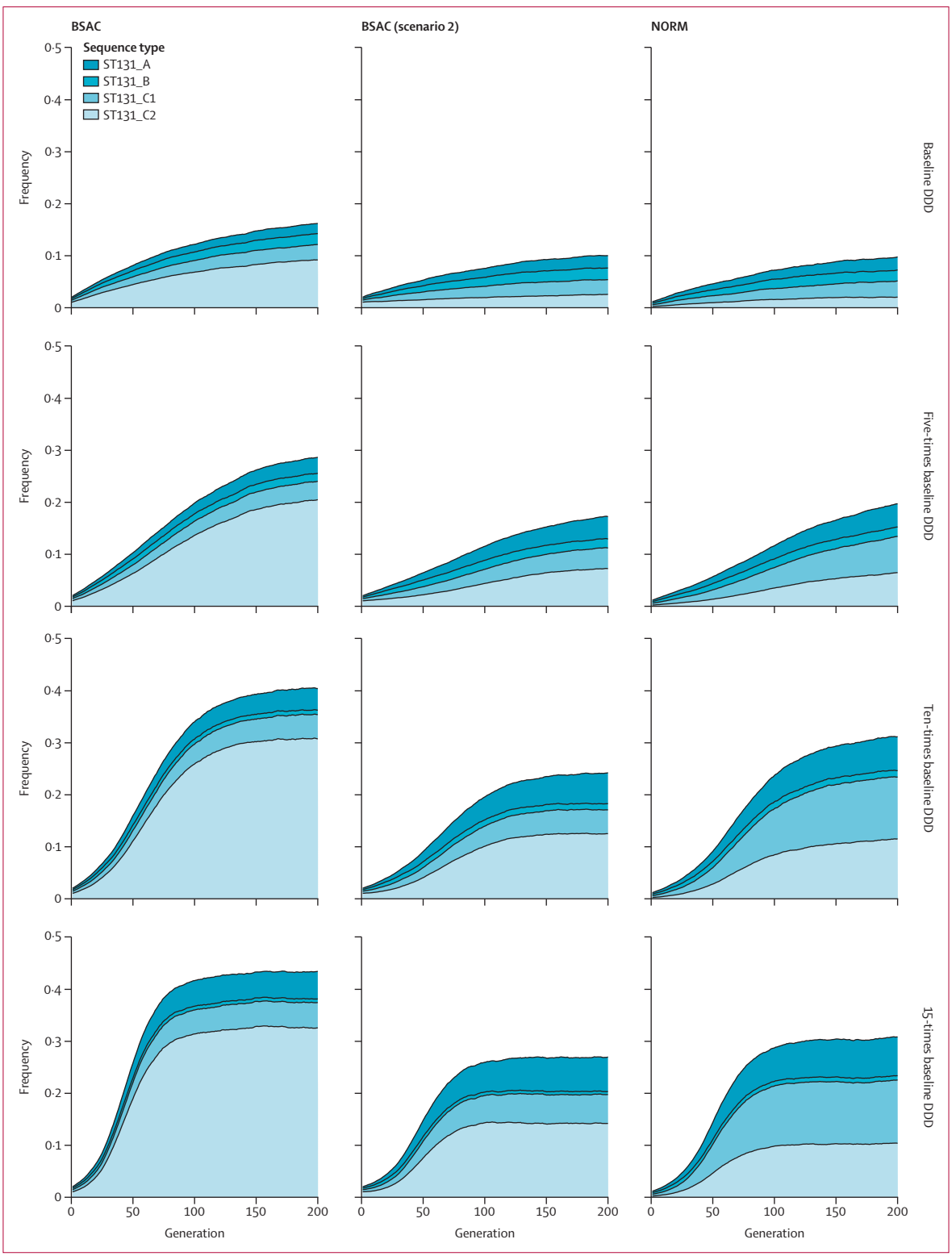

Figure 5: **Mean relative frequency changes for ST131 clades A, B, C1, and C2 over time under increasing DDD selection levels**
BSAC and NORM evolve towards their respective empirical equilibrium frequencies. BSAC (scenario 2) evolves towards the NORM empirical equilibrium frequencies, although the actual strains used to set up the simulation are the same as BSAC. Each scenario is subject to different DDD selection levels. The actual simulation includes all clones (appendix 2 p 11), but only ST131 clades are shown here. DDD=defined daily doses. BSAC=British Society for Antimicrobial Chemotherapy. NORM=Norwegian Surveillance Programme on Resistant Microbes. ST=sequence type.

operate in the same manner in different successful extant or expanding clones.

A large genomic survey in the USA in 2019–20 found the same dominant clones among predominantly urine isolates of *E coli* as observed in bloodstream infection cases in the UK and Norway.[19] It is more challenging to extrapolate the relative effects seen in the current study to regimes of DDD levels observed in regions with much less regulated and also non-documented use of antibiotics. As examples, a study of healthy colonisation dynamics conducted in Laos found no isolates of ST131 among 219 ESBL-positive *E coli*.[20] Similarly, screening of patients in intensive care units at two hospitals in Hanoi, Viet Nam, identified ST131 only as the fourth most common clone, with ST648, ST410, and ST617 at higher frequencies.[21] Despite ST131 being recognised as the most successful MDR extra-intestinal pathogenic *E coli* clone globally, it might not have the same selective advantage among *E coli* circulating in southeast Asia, compared with Europe and the USA.[8] Data on documented antibiotic use from the WHO GLASS surveillance network support this interpretation, because use of non-penicillin β-lactam antibiotics is substantially higher in countries within this region, such as Laos, Mongolia, and Nepal,[21] than in the UK. Taken together, findings from our study and the other available data show that the effects of antibiotic selection and balancing selection stemming from non-antibiotic-related factors in colonisation competition might be markedly different in ultra-selective environments, warranting sustained research effort in quantifying the drivers of success for clinically important pathogens across a range of ecological settings.

Contributors

JC conceptualised, designed, and obtained funding for the study. JP obtained additional funding for the study. AKP, RAG, HP, and JC developed the methods. AKP and FC prepared samples for DNA extraction. AKP and RAG performed genomic analysis. BJP reviewed and provided data from the BSAC collection. AKP, RAG, HP, MP, and JC analysed, visualised, and interpreted the data. JC had the main responsibility for drafting the manuscript, with contributions from AKP, RAG, and HP. JP, TK, GSS, NJC, AM, PJJ, and ØS reviewed and interpreted the data and results. All authors reviewed and edited the manuscript. All authors had full access to and verified all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

NJC reports support to his institution from GlaxoSmithKline, Pfizer, and Merck (MSD), and personal grants from MRC, outside the submitted work. All other authors declare no competing interests.

Data sharing

Sequence data on the BSAC2 collection are publicly available on the European Nucleotide Archive within project PRJEB44839. Source code for running the simulation-based analyses is available within a public Github repository (https://github.com/hpesonen/MDR_NFDS_population_simulation).

Acknowledgments

This work was supported by the Trond Mohn Foundation (antimicrobial resistance grant TMS2019TMT04 to AKP, RAG, JC, ØS, and PJJ), Marie Skłodowska–Curie Actions (801133 to AKP), the European Research Council (grant 742158 to JC and HP), and the Wellcome Trust and Royal Society (grant 104169/Z/14/A to NJC). Sequencing at the Wellcome Sanger Institute was supported by a Wellcome Trust grant (098051).

References

1 Denamur E, Clermont O, Bonacorsi S, Gordon D. The population genetics of pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2021; **19:** 37–54.
2 Bonten M, Johnson JR, van den Biggelaar AHJ, et al. Epidemiology of *Escherichia coli* bacteremia: a systematic literature review. *Clin Infect Dis* 2021; **72:** 1211–19.
3 Goossens H, Ferech M, Vander Stichele R, Elseviers M. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. *Lancet* 2005; **365:** 579–87.
4 Gladstone RA, McNally A, Pöntinen AK, et al. Emergence and dissemination of antimicrobial resistance in *Escherichia coli* causing bloodstream infections in Norway in 2002–17: a nationwide, longitudinal, microbial population genomic study. *Lancet Microbe* 2021; **2:** e331–41.
5 Kallonen T, Brodrick HJ, Harris SR, et al. Systematic longitudinal survey of invasive *Escherichia coli* in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. *Genome Res* 2017; **27:** 1437–49.
6 Marin J, Clermont O, Royer G, et al. The population genomics of increased virulence and antibiotic resistance in human commensal *Escherichia coli* over 30 years in France. *Appl Environ Microbiol* 2022; **88:** e0066422.
7 Lipworth S, Vihta K-D, Chau K, et al. Ten-year longitudinal molecular epidemiology study of *Escherichia coli* and *Klebsiella* species bloodstream infections in Oxfordshire, UK. *Genome Med* 2021; **13:** 144.
8 Price LB, Johnson JR, Aziz M, et al. The epidemic of extended-spectrum-β-lactamase-producing *Escherichia coli* ST131 is driven by a single highly pathogenic subclone, H30-Rx. *MBio* 2013; **4:** e00377-13.
9 McNally A, Kallonen T, Connor C, et al. Diversification of colonization factors in a multidrug-resistant *Escherichia coli* lineage evolving under negative frequency-dependent selection. *MBio* 2019; **10:** e00644-19.
10 Page AJ, De Silva N, Hunt M, et al. Robust high-throughput prokaryote *de novo* assembly and improvement pipeline for Illumina data. *Microb Genom* 2016; **2:** e000083.
11 Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; **18:** 821–29.
12 Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014; **30:** 2068–69.
13 Roer L, Tchesnokova V, Allesøe R, et al. Development of a web tool for *Escherichia coli* subtyping based on *fimH* alleles. *J Clin Microbiol* 2017; **55:** 2538–43.
14 Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, et al. Sequential acquisition of virulence and fluoroquinolone resistance has shaped the evolution of *Escherichia coli* ST131. *MBio* 2016; **7:** e00347-16.
15 Brolund A, Sundqvist M, Kahlmeter G, Grape M. Molecular characterisation of trimethoprim resistance in *Escherichia coli* and *Klebsiella pneumoniae* during a two year intervention on trimethoprim use. *PLoS One* 2010; **5:** e9233.
16 European Centre for Disease Prevention and Control. Latest surveillance data on antimicrobial consumption. 2022. https://www.ecdc.europa.eu/en/antimicrobial-consumption/surveillance-and-disease-data/database (accessed June 13, 2023).
17 Mäklin T, Thorpe HA, Pöntinen AK, et al. Strong pathogen competition in neonatal gut colonisation. *Nat Commun* 2022; **13:** 7417.
18 Sundqvist M, Geli P, Andersson DI, et al. Little evidence for reversibility of trimethoprim resistance after a drastic reduction in trimethoprim use. *J Antimicrob Chemother* 2010; **65:** 350–60.
19 Mills EG, Martin MJ, Luo TL, et al. A one-year genomic investigation of *Escherichia coli* epidemiology and nosocomial spread at a large US healthcare network. *Genome Med* 2022; **14:** 147.
20 Kantele A, Kuenzli E, Dunn SJ, et al. Dynamics of intestinal multidrug-resistant bacteria colonisation contracted by visitors to a high-endemic setting: a prospective, daily, real-time sampling study. *Lancet Microbe* 2021; **2:** e151–58.
21 WHO. Global antimicrobial resistance and use surveillance system (GLASS) report: 2022. Dec 9, 2022. https://www.who.int/publications/i/item/9789240062702 (accessed June 13, 2023).