The Identification of Sources and Genomic Characteristics of Pandemic Extraintestinal Pathogenic *Escherichia coli*

By

Yuan Hu Allegretti

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Epidemiology
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:

Professor Lee W. Riley, Co-chair
Professor John Colford, Co-chair
Professor Jay Graham
Professor Arthur Reingold

Fall 2021

Abstract

The Identification of Sources and Genomic Characteristics of Pandemic Extraintestinal Pathogenic *Escherichia coli*

By

Yuan Hu Allegretti

Doctor of Philosophy in Epidemiology
University of California Berkeley
Professor Lee W. Riley, Co-chair
Professor John Colford, Co-chair

This dissertation investigates the epidemiologic and microbiologic features of extraintestinal pathogenic *Escherichia coli* (ExPEC) organisms, with a particular emphasis on antimicrobial drug resistance (AMR) associated with these ExPECs. The first section and the second section (Chapter 2 and Chapter 3) describes the risk factors for healthy adults to carry drug-resistant *E. coli* in their gut. Chapter 2 presents an overview of what has been described in the literature on risk factors for fecal carriage of drug-resistant *E. coli*. Chapter 3 focuses on a study we conducted on campus here at the University of California, Berkeley to examine risk factors for carriage of drug-resistant *E. coli* using a penalized regression model. Chapter 4 describes findings based on molecular epidemiologic and computational biologic methods to investigate genetic features associated with the lack of carriage of drug-resistance genes in a major pandemic lineage of ExPEC, ST95. The focus on a lineage that lacks drug-resistance genes was an attempt to characterize counterfactual genetic factors associated with AMR ExPECs that may help to better understand why the prevalence of AMR infections are caused by a limited set of ExPEC lineages. We were able to determine that certain dietary behavior and travel destinations can contribute to colonization with drug-resistant *E. coli*. Such knowledge could be potentially used to devise prevention of acquisition of drug-resistant *E. coli*. We also found potential bacterial mechanisms to resist acquiring drug resistant genes. Such knowledge could be exploited to devise biologic interventions to interrupt *E. coli* from gaining drug-resistance genes. In conclusion, this dissertation work highlights the importance of global bacterial drug resistant infectious disease problem and opened a path towards further studies that may ultimately lead to the creation of new prevention strategies against AMR infections.

List of Tables

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 2

Chapter 3

Chapter 4

Acknowledgements

First, I would like to show my deepest gratitude to my advisor and mentor, Dr. Lee Riley, for his continuous support in my academic study. Thank you for letting me pursue everything I was interested in. Your advice was the most valuable education I have ever received to grow not only as a scientist but also as a person.

I would also like to show my gratitude to my dissertation committees, Dr. John Colford, Dr. Art Reingold, and Dr. Jay Graham for your generous support and valuable comments in my dissertation studies.

Thank you to all of Riley's lab members who worked on the projects together and gave me valuable comments. Thank you Dr. Sheila Adams-Sapper for training me when I joined the lab in 2017. Thank you Dr. Reina Yamaji, Dr. Yusuke Matsui, Ms. Julia Rubin, and Dr. Reginara Assis for helping me achieve the projects necessary for this dissertation. And thank you Ms. Cheyenne Butcher for your continuous support as a co-worker and as a friend.

I would also like to thank my advisors in Japan for making my journey to the United States possible. Dr. Kazuhisa Sekimizu, thank you for teaching me how to be a scientist. Dr. Kiyoshi Kita and Dr. Takafumi Tsuboi, thank you for introducing me to the world of global health.

I would also like to thank my parents, Dr. Bai Hu and Rongjiu Zou for always encouraging my dream. I cannot show how grateful I am to you for always believing in me. Thank you for showing me efforts will absolutely not betray me.

Lastly, I would like to thank my husband, Dr. Dylan Allegretti, for your emotional support and advice as a scientist and as a person. Your passion for math and perseverance has always motivated me and made my years in Berkeley more meaningful and happier.

Table of Contents

Chapter 1. Introduction

1.1 ExPEC as foodborne pathogens

Escherichia coli (E. coli) is divided into three groups—1) those that reside in mammalian intestine as commensal bacteria, 2) those that cause diarrhea, referred to as intestinal pathogenic *E. coli* (IPEC), and 3) those that cause extraintestinal infections, referred to as extraintestinal pathogenic *E. coli* (ExPEC). IPEC is divided into 6 well-known pathotypes: enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EAEC), diffusely adherent *E. coli* (DAEC) and enterotoxigenic *E. coli* (ETEC)(1). ExPEC is the most common cause of bloodstream infections (BSI) or sepsis(3) and urinary tract infections (UTI)(4, 5), as well as neonatal meningitis. They represent a major health threat, especially when they become resistant to antimicrobial drugs (6). BSIs are associated with high morbidity and mortality and are frequently preceded by an episode of UTI(3). UTIs are estimated to affect 150 million people each year worldwide, and contribute to severe economic loss (7).

Based on recent molecular epidemiologic studies, it has become evident that nearly half of all ExPEC infections in humans in any community or institutional settings are caused by a limited set of lineages or sequence types--ST69, ST73, ST95, ST131, and ST393--based on multilocus sequence typing (MLST)(4, 8). These prevalent genotypes are also known as pandemic or intercontinental ExPEC lineages(4). More recently, additional 13 STs (ST405, ST38, ST648, ST410, ST354, ST12, ST167, ST58, ST617, ST88, ST23, ST117, and ST1193) have been described that have been reported by more than 28 to 85 studies globally in a meta-analysis in 2019(9). Multiple recent studies suggest that UTI is caused by ExPEC in gut microbiota(10) and that food may be an important reservoir of ExPEC(11–13). A recent study conducted at a college community found that more than 50% of the ExPEC genotypes were shared by isolates from urine samples from patients with UTI and from fecal samples obtained from healthy volunteers (14). The same study found that the prevalence of the dominant genotypes isolated from healthy adults' fecal samples fluctuated by months in the 6 months study period. A prospective cohort study reported that vegetarian diet was associated with a decreased risk of UTI(15). These findings suggest that ExPEC may have reservoirs outside of the human intestine and that they are not members of the intestinal commensal *E. coli* that breach a sterile barrier to cause disease.

This dissertation is divided into three sections that explore epidemiologic and microbiologic features of ExPEC organisms, with a particular emphasis on antimicrobial drug resistance (AMR) associated with these ExPECs. *E. coli* organisms serve as a major reservoir of mobile DNA elements carrying drug-resistance genes that are spread among populations of other Gram-negative bacteria (GNB). Thus, ExPECs serve as a "gateway" organism for many types of AMR GNB-associated infections. By studying ExPECs, we may be able to better understand the

epidemiologic features of other AMR GNB pathogens.

The first section and the second section (Chapter 2 and Chapter 3) describe the risk factors for healthy adults to carry drug-resistant *E. coli* in their gut. Chapter 2 presents an overview of what has been described in the literature on risk factors for fecal carriage of drug-resistant *E. coli*. Chapter 3 focuses on a study we conducted on campus here at the University of California, Berkeley to examine risk factors for carriage of drug-resistant *E. coli* using a penalized regression model. Finally, Chapter 4 describes findings based on molecular epidemiologic and computational biologic methods to investigate genetic features associated with the lack of carriage of drug-resistance genes in a major pandemic lineage of ExPEC, ST95. The focus on a lineage that lacks drug-resistance genes was an attempt to characterize counterfactual genetic factors associated with AMR ExPECs that may help to better understand why the prevalence of AMR infections are caused by a limited set of ExPEC lineages.

1.2 The risk factors for intestinal carriage of drug-resistant *E. coli*

Antimicrobial resistance is one of the most pressing public health challenges of our time. In particular, the rising incidence of infections caused by drug-resistant Gram-negative bacteria is a serious problem due to the potential for rapid spread of resistance via mobile elements and limited treatment options (16–18).

Among Gram-negative bacteria developing drug resistance, *E. coli* is the most frequent cause of extraintestinal infections such as urinary tract infection and bloodstream infection (17). Drug-resistant intestinal pathogenic *E. coli*, such as Shiga toxin-producing *E. coli* (STEC), are also increasingly recognized (19, 20). *E. coli* can be transmitted through contaminated water or food, or through contact with people and other animals (21). The prevalence and incidence of infections caused by drug-resistant pathogenic *E. coli* have been rapidly increasing worldwide (17, 22, 23).

Major sources of drug-resistant bacteria include the environment such as contaminated water (24), food including meat (25, 26) and vegetables (27, 28), and healthcare settings (29). Additionally, intestinal commensal drug-resistant bacteria have been reported as an important reservoir of antimicrobial drug resistance genes (ARGs) (30, 31). Surveillance on human fecal carriage of drug-resistant bacteria has revealed that there is an increasing trend in intestinal ARG carriage worldwide (22, 32). Numbers of studies have independently reported potential risk factors for the intestinal carriage of drug-resistant bacteria. Most of these studies have found previous antibiotic use to be associated with drug-resistant bacteria carriage in both primary care patients and healthy populations (33, 34). Also, traveling to developing countries has been identified as a risk factor for acquiring drug-resistant bacteria (35). Risk factors related to healthcare-associated infections (HAI) have been reported as well, including admission to the intensive care unit (ICU), use of catheter, and dialysis (36–38).

E. coli is also a member of the commensal flora of human and other mammalian animal intestinal tracts. As such, they can acquire ARGs by horizontal gene transfer (39) from drug-resistant *E. coli* strains and other Gram-negative bacteria that enter the intestinal tract via exposures to contaminated food, water, and other external sources. Thus, risk factors for fecal carriage of drug-resistant commensal *E. coli* and ARGs could include exposures to environmental sources of drug-resistant bacteria, in addition to traditional risks such as prior use of antibiotics.

The impact or magnitude of exposures to food on the commensal *E. coli* carriage of ARGs is not known. Identifying risk factors for fecal carriage of drug-resistant commensal *E. coli* associated with food could potentially improve public health intervention to prevent the spread of drug-resistant *E. coli* and ARGs. While a recent review studied risk factors for fecal carriage of Gram-negative bacteria expressing extended-spectrum beta-lactamase (ESBL) reported by papers from OECD countries from 1978 to 2015(34), there has not been a comprehensive analysis of more recent literature reporting other resistance mechanisms of human commensal *E. coli* nor surveillance focusing on the relationship of dietary habit and ARG carriage.

1.3 The genomic features of pan-susceptible ST95

Almost half of ExPEC infections are caused by pandemic lineages, ST69, ST73, ST95, ST131, and ST393 (4). More than 70% of ExPEC infections of these pandemic lineages are caused by drug resistant strains, which is a serious public health problem leading to reduced efficiency of current antibiotics. Particularly, ST131 is associated with drug resistance not only in the US but also worldwide (44–46). ST95, however, remain pan-susceptible to most of the antimicrobial drugs today (4, 5), despite the selective pressure to acquire resistance due to the excessive use of antimicrobial drugs in hospitals, food animal husbandry, and agriculture. This makes ST95 unique from other dominant ExPEC lineages. Mechanisms for ST95 remaining pan-susceptible to drugs, or resistance to drug resistance acquisition, remain unknown.

Previous studies have identified several genomic features associated with pan-susceptibility of ST95. Among ST95 isolates, the fimH6 lineage and the presence of UTI89 like plasmid (pUTI89*), homologs of the 114-kb IncFIB/IncFII plasmid found on UTI89 (pUTI89, NCBI accession number CP000244), was significantly associated with pan-susceptibility to antibiotics (5). Both of these factors were also associated with reduced carriage of other plasmids. pUTI89 carries many putative virulence plasmids, and is reported to be important for early stage of bacterial infection (47). We hypothesize that carriage of pUTI89* inhibits ST95 from acquiring other plasmids or mobile genetic elements that carry drug-resistance genes, which prevents ST95 from acquiring antimicrobial drug resistance genes (ARGs).

We will focus on the pangenome of our ST95 collection. Microbial genomes can be divided into core genome that contains genes present in all strains and accessory genome that contains genes

present in two or more strains, or unique genes found in only one strain (48). Pangenome of a bacterial species is the entire set of genes comprised of a combination of core genes and all accessory genes from each isolate (48). Excess use of antibiotics is known to raise the general rates of mutation, recombination, and gene transfer in all the microbiome, and simultaneously providing the selective force to recruit more genes into microbial genomes (49), leading to larger size of the pangenome. It is also reported that pangenome analysis will yield a higher resolution in genetic variation even within one MLST genotype, which is merely based on 7 genes in the core genomes (50). Preliminary search of the whole genome sequences of our ST95 collection (n=106) revealed that the pan-genome consists of 11037 genes, of which 3399 are core genes, and that there is a high variability in the number of genes on each ST95 isolate. Therefore, our hypothesis is that genomic features on accessory genomes as well as on core genomes or plasmids are associated with pan-susceptibility of ST95. We will aim to determine the association between the presence or absence of accessory genes and drug pan-susceptibility.

1.4. Summary

Our purpose is to investigate risk factors associated with intestinal carriage of drug-resistant commensal *E. coli*. We also aim to identify risk factors related to food. To answer this question, we will conduct literature search for the recent five years and a small-scale surveillance in a college community in Chapter 2. In the literature search, we will focus on the recent five years because of the increasing prevalence of multiple mechanisms of resistance among Gram-negative bacteria causing extraintestinal and intestinal infections during this period, including mechanisms such as ESBL (40, 41), carbapenemase (42), and metallo-beta-lactamase production (41), and plasmid-mediated colistin resistance (43). In Chapter 3, we will focus on the association between food or food-related factors and the risk of carrying drug-resistant *E. coli* or AMR genes by the surveillance in a college community.

Chapter 2: Risk factors for fecal carriage of drug-resistant *Escherichia coli*: a systematic review and meta-analysis

A version of this chapter was published in the Antimicrobial Resistance & Infection Control in December 2020.
Reference: Hu, Y., Matsui, Y. & Riley, L. W. Risk factors for fecal carriage of drug-resistant *Escherichia coli*: a systematic review and meta-analysis. Antimicrob Resist Infect Control 9, 31 (2020). https://doi.org/10.1186/s13756-020-0691-3.

Introduction

Antimicrobial resistance is one of the most pressing public health challenges of our time. In particular, the rising incidence of infections caused by drug-resistant Gram-negative bacteria is a serious problem due to the potential for rapid spread of resistance via mobile elements and limited treatment option (17, 18, 51).

Among Gram-negative bacteria developing drug resistance, *Escherichia coli* (*E. coli*) is the most frequent cause of extraintestinal infections such as urinary tract infection and bloodstream infection (17). Drug-resistant intestinal pathogenic *E. coli*, such as Shiga toxin-producing *E. coli* (STEC), are also increasingly recognized (20, 52). *E. coli* can be transmitted through contaminated water or food, or through contact with people and other animals (21). The prevalence and incidence of infections caused by drug-resistant pathogenic *E. coli* have been rapidly increasing worldwide (17, 22, 23).

Major sources of drug-resistant bacteria include the environment such as contaminated water (24), food, including meat (25, 26) and vegetables (28, 53), and healthcare settings (29). Additionally, intestinal commensal drug-resistant bacteria have been reported as an important reservoir of antimicrobial drug resistance genes (ARGs) (54, 55). Surveillance on human fecal carriage of drug-resistant bacteria has revealed that there is an increasing trend in intestinal ARG carriage worldwide (22, 32).

Numbers of studies have independently reported potential risk factors for the intestinal carriage of drug-resistant bacteria. Most of these studies have found previous antibiotic use to be associated with drug-resistant bacteria carriage in both primary care patients and healthy populations (33, 34). Also, traveling to developing countries has been identified as a risk factor for acquiring drug-resistant bacteria (35). Risk factors related to healthcare-associated infections (HAI) have been reported as well, including admission to the intensive care unit (ICU), use of catheter, and dialysis (36–38).

E. coli is also a member of the commensal flora of human and other warm-blooded animal intestinal tracts. As such, they can acquire ARGs by horizontal gene transfer (39) from drug-resistant *E. coli* strains and other Gram-negative bacteria that enter the intestinal tract via exposures to contaminated food, water, and other external sources. Thus, risk factors for fecal carriage of drug-resistant commensal *E. coli* and ARGs could include exposures to environmental sources of drug-resistant bacteria, in addition to traditional risks such as prior use of antibiotics.

The impact or magnitude of exposures to food on the commensal *E. coli* carriage of ARGs is not known. Identifying risk factors for fecal carriage of drug-resistant commensal *E. coli* associated with food could potentially improve public health intervention to prevent the spread of drug-resistant *E. coli* and ARGs. While a recent review studied risk factors for fecal carriage of Gram-negative bacteria expressing extended-spectrum beta-lactamase (ESBL) reported by papers from OECD countries from 1978 to 2015 (34), there has not been a comprehensive analysis of more recent literature reporting other resistance mechanisms of human commensal *E. coli*.

The purpose of this review was to investigate risk factors associated with intestinal carriage of drug-resistant commensal *E. coli* in the recent five years. We also aimed to identify risk factors related to food. We focused on the recent five years because of the increasing prevalence of multiple mechanisms of resistance among Gram-negative bacteria causing extraintestinal and intestinal infections during this period, including mechanisms such as ESBL (41, 56), carbapenemase (42), and metallo-beta-lactamase production (41), and plasmid-mediated colistin resistance (43).

Methods

Data Sources and Search Strategy

The protocol of this meta-analysis was not preregistered. We performed a systematic review and meta-analysis following the PRISMA [Preferred Reporting Items for Systematic Reviews and Meta Analyses] guidelines (57) (Supplementary Table 1). We conducted a literature search with the databases PubMed, Embase, and Web of Science. We limited the search to articles published between 2014 and 2019. Only articles published in English were included.

The search focused on risk factors for intestinal carriages of drug-resistant commensal *E. coli*, which was conducted on August 9th, 2019. For the purpose of this review, the definition of antimicrobial drug resistance was based on the drug-susceptibility test results (disk diffusion test, minimum inhibitory concentration (MIC) test, VITEK) reported by the clinical microbiology or research laboratories described in the reviewed studies, which followed the guidelines of organizations such as the Clinical and Laboratory Standards Institute (CLSI). We included reports of *E. coli* resistance to beta-lactams, aminoglycosides, fluoroquinolones, and tetracyclines. We included search terms: (feces [Title/Abstract] OR stool [Title/Abstract] OR fecal [Title/Abstract] OR faecal [Title/Abstract] OR "rectal swab" [Title/Abstract]) AND ("escherichia coli" [Title/Abstract] OR escherichia [Title/Abstract] OR "e.coli" [Title/Abstract])) AND ("drug resistant" [Title/Abstract] OR "drug susceptible" [Title/Abstract] OR "drug susceptibility" [Title/Abstract] OR "antimicrobial resistance" [Title/Abstract] OR "antimicrobial resistant" [Title/Abstract] OR resistant [Title/Abstract] OR resistance [Title/Abstract] OR drug [Title/Abstract] OR multidrug [Title/Abstract]) AND (questionnaire [Title/Abstract] OR surveillance [Title/Abstract] OR survey [Title/Abstract]).

After the databases were reviewed, the results were exported and then compiled with the reference management software Covidence (58). Duplicates were removed by automated process of Covidence, followed by a manual search to identify and remove additional duplicates.

Study Selection

All abstracts were screened first by author YH and then by author YM to minimize omission of eligible studies. Screening criteria were as follows: (1) examined bacteria must include *E. coli* or Enterobacteriaceae; (2) examined bacteria must be isolated from human feces, stool, or rectal swab; (3) must report risk factors.

Studies reporting risk factors for drug-resistant Enterobacteriaceae were considered eligible because *E. coli* is the most common Enterobacteriaceae.

Studies that remained of interest were then screened based on their full text by two independent reviewers, YH and YM. Disagreements were resolved by consensus.

Inclusion criteria were: (1) reported risk factor(s); (2) reported measure of associations and

accompanying 95% confidence intervals (95% CI) or its equivalent; (3) study population aged 18-65; (4) healthy study population; (5) survey conducted after 2010.
For the meta-analysis, we excluded studies that (1) did not report risk factors commonly assessed in 3 or more studies or (2) did not offer sufficient data to create a contingency table.

Data Extraction
Data were first extracted by YH and checked by YM. The assessment measures extracted from the included studies were as follows: (1) publication data: lead author names, year of publication; (2) demographic and epidemiological data: location of study, study population, study design, sample size, outcome, prevalence of drug-resistant bacteria, outcome measurement methods, statistical analysis methods; (3) risk-factor associated data: risk factor(s) investigated, measure of associations (odds ratios, risk ratios or prevalence ratios) and accompanying 95% CI.
When enumerating risk factors from each eligible study, we did not limit the analysis to statistically significant factors to avoid publication bias and to identify as many factors studied to date as possible.

Meta-analysis
For studies which provided enough data to allow for the creation of contingency tables, unless the authors reported an adjusted OR and corresponding 95% CI, we manually calculated the OR and 95% CI. If there were insufficient data to create a contingency table, we excluded the study to calculate pooled estimates.
We performed random-effects meta-analysis under a Mantel-Haenszel model with Hartung-Knapp adjustment to estimate the pooled effect of each commonly reported risk factors for intestinal carriage of drug-resistant *E. coli*. Mantel-Haenszel random-effects model estimates the amount of between-study variation by comparing each study's result with a fixed-effect meta-analysis result but avoids approximating Normal distributions (59, 60). Hartung-Knapp adjustment provides a more conservative and robust pooled OR estimates and 95% CI, allowing for any heterogeneity between studies even when the study number is small and study size is unequal (61). Forest plots were created to visualize the reported OR and 95% CI from each studies and pooled ORs for each commonly assessed risk factors. We assessed statistical heterogeneity between studies by the Chi^2 test and variation due to heterogeneity across the studies by the I^2 statistic. $P<0.10$ was considered indicative of statistically significant heterogeneity in the Chi^2 test, and I^2 values of 25, 50 and 75% were defined as low, moderate, and high estimates, respectively. We evaluated the potential for publication bias with funnel plots and Egger's tests for meta-analyses with at least 10 studies (62), which test for asymmetry

of the funnel plot and effects of small studies. Analyses were conducted with R version 3.5.1 (63), with package 'meta' version 4.9-6 (64).

Results

Study selection

Our search identified 395 unique studies that we assessed for eligibility with title and abstract screening. Of these, 58 studies were forwarded to full-text article screening. Of the 58 full-text articles, we identified 15 relevant articles that reported risk factors associated with drug-resistant Enterobacteriaceae or *E. coli* carriage (65–79).

Twelve of 15 studies included in the systematic review were eligible for inclusion in the meta-analysis, which reported sufficient data to create contingency tables to compare risk factors that were studied in at least three of the studies (65–68, 70, 71, 74–79). See Figure 1, Table 1, and Supplementary Table 2 for further details of search and reasons for exclusion.

Study characteristics

The 15 studies represented 8 countries: England, Gambia, Germany, Netherlands, Northern Cyprus, Singapore, Sweden, and Tanzania (Table 1). None of the studies reported randomization in participant selection. Eight studies sampled volunteers from healthy general population that were registered to a hospital system. Five were cohort studies of healthy travellers that compared the prevalence of drug-resistant Enterobacteriaceae or *E. coli* before and after the travel. Two studies surveyed pig farmers.

Five studies reported prevalence of drug-resistant *E. coli*, while 10 studies investigated Enterobacteriaceae. The frequency of *E. coli* among Enterobacteriaceae ranged from 79-97% for 9 studies, while one study reported 29%. All studies collected information on demographic factors, behaviors, and past illness from participants. Some studies excluded insufficient response from the surveys.

The prevalence of fecal drug-resistant Enterobacteriaceae reported in the studies ranged from 1% to 51%. The pooled prevalence was 14% (95% CI 8-23%) (Figure 2A). Nine studies reported ESBL producing Enterobacteriaceae. The pooled prevalence of ESBL-producing Enterobacteriaceae was 18% (95%CI 9-31%) (Figure 2B). The prevalence among general population was 8% (95%CI 4-14%) (Figure 2C) and among travellers was 37% (95%CI 30-43%) (Figure 2D). All studies followed established drug susceptibility testing methods, disc diffusion tests, VITEK 2, or minimum inhibitory concentration (MIC) measurement. Common statistical methods for risk factor analysis included univariate and multivariate logistic regression, chi-squared test, and Fisher's exact *t* test.

Commonly assessed risk factors

Commonly assessed risk factors identified in this review are shown in Table 2. We identified fourteen risk factors assessed in three or more studies. We assessed the pooled ORs in the meta-

analysis (Table 2, Figure 3A, Supplementary Figure 1A).

Traveling to India was the only risk factor that all studies reported to be significantly associated with fecal carriage of drug-resistant *E. coli*. For the remaining risk factors, ORs and accompanying 95% CI were found to vary among studies. There were three risk factors that showed significant pooled ORs. These included antimicrobial use within the previous 12 months (OR 1.84 [95% CI 1.35-2.51]), diarrhea symptoms (OR 1.56 [95% CI 1.09-2.25]), and vegetarian diet (OR 1.60 [95% CI 1.00(1.0043)-2.56(2.5587)]). Six (46%) of 13 studies found antimicrobial use in the previous 12 months, 4 (57%) of 7 studies found diarrhea symptom, and 2 (40%) of 5 studies found vegetarian diet to be significantly associated with the carriage of drug-resistant bacteria.

Smoking, living with pet(s), gender, education level, previous hospital admission, proton-pump inhibitor (PPI) use, chronic disease, international travel, travel to Southeast Asia and exposure to livestock were commonly assessed but no significant pooled OR was found in these studies. Of these commonly assessed risk factors, three factors (PPI use, chronic disease, travel to Southeast Asia) were reported as significant risks among half or more studies included in this review. Two (67%) of 3 studies found PPI use, 2 (67%) of 3 studies found chronic disease, and 4 (50%) of 8 studies found travel to Southeast Asia to be significantly associated with the carriage of drug-resistant bacteria.

Risk factors based on travelling status

The prevalence of drug-resistant *E. coli* carriage suggested two distinct populations. We divided the population into travellers and other general population adults and replicated the analysis (Table 3, Figure 3B, 3C, Supplementary Figure 1B, 1C).

Antimicrobial use within the previous 12 months, diarrhea symptoms, gender, travelling to India, travelling to Africa, and travelling to Southeast Asia were assessed for travellers. We also assessed antimicrobial use within the previous 12 months, diarrhea symptoms, gender, travelling abroad, travelling to Southeast Asia, education status, pet, and previous hospitalization among general population adults. The results showed that antimicrobial use within the previous 12 months (OR 2.81 [95% CI 1.47-5.36]), diarrhea symptoms (OR 1.65 [95% CI 1.02-2.68]), vegetarian diet (OR 1.92 [95% CI 1.13-3.26]), and travelling to India (OR 3.80 [95% CI 2.23-6.47]) remained significant risk factors among travellers. Among general population adults, antimicrobial use within the previous 12 months (OR 1.51 [95% CI 1.17-1.94]), diarrhea symptoms (OR 1.53 [95% CI 1.27-1.84]), and travelling to Southeast Asia (OR 1.67 [95% CI 1.02-2.73]) were significant risk factors.

Risk factors related to food

Six of 15 studies reported risk factors related to food. Five studies assessed the risk among vegetarians (Table 2). As stated above, pooled OR showed significant association with being a vegetarian (OR 1.60 [95% CI 1.00-2.56]). Two studies reported significant association, one with unadjusted OR (65), and another with adjusted OR (77).

Four studies reported potential food-associated risk factors other than being a vegetarian. One study reported exposure to raw milk as significant risk factor for acquiring multi-drug-resistant *E. coli* (OR 7.54 [95% CI 2.41-23.45]) (72). Two studies reported the effect of eating street food during travel. One of them was reported as significant risk (OR 2.09 [95% CI 1.30-3.38] for daily consumption; OR was 1.37 [95% CI 1.08-1.73] for occasional consumption during travel) (65). Another study did not find significant association (OR 0.92 [95\% CI 0.49-1.74]) (75). Two studies assessed the effect of raw vegetable consumption on the fecal carriage of drug-resistant *E. coli*. One of them reported that raw vegetable consumption during a trip to Southeast Asia significantly increased the risk of intestinal carriage of drug-resistant Enterobacteriaceae (OR 2.18 [95% CI 1.29-3.68]), while exposure to raw vegetable in South Asia significantly decreased the risk (OR 0.34 [95% CI 0.12-0.93]) (65). The other study did not find any significant association (OR 0.58 [95% CI 0.33-1.07]) (76).

Bias assessment and heterogeneity evaluation

We evaluated heterogeneity among studies, and potential extent of publication bias in meta-analysis (Table 2, Table 3, Figure 4, Figure3B, 3C). Funnel plots of all studies reporting significant association (Figure 4) were generated to assess the potential extent of publication bias.

For pooled estimates of all studies, risk factors related to travel showed high chi^2 (11-81, $P<0.01$) and I^2 value (53-94%) except for travel to India. This suggests that there was substantially high heterogeneity among studies that examined the effect of international travel, travel to Southeast Asia, and travel to Africa, respectively. Smoking, PPI use, and chronic disease status also showed moderate to high heterogeneity (I^2 66-77%). For all other risk factors, no heterogeneity was observed, suggesting that the evidence was of high quality.\\

For stratified estimates among travellers, travel to Africa and travel to Southeast Asia were the only risk factors that showed high heterogeneity (chi^2 19.27 and 41.24, respectively, $p<0.01$, and I^2 90%). Among general adults, travel abroad and travel to Southeast Asia showed moderate heterogeneity (chi^2 10.73 and 5.56, respectively, $p=0.06$, and I^2 53-64%).

The shapes of the funnel plots were approximately symmetrical for significant risk factors, and Egger's test showed P = 0.42 for antimicrobial use within the previous 12 months among all populations included in this study (Figure 4). This suggests that no publication bias existed for this factor. For all other risk factors, due to the insufficient number of studies (less than 10 studies for each), we did not evaluate the potential for publication bias with funnel plots and Egger's tests

for small study effects (62).

Discussions

This study summarizes risk factors associated with intestinal carriage of drug-resistant Enterobacteriaceae, in particular, *E. coli* among healthy adults. Our systematic review and meta-analysis on studies published from 2014 to 2019 identified several risk factors for intestinal carriage of drug-resistant *E. coli*. We found evidence for our hypothesis that commensal *E. coli* can acquire ARGs carried by Gram-negative bacteria that enter the intestinal tract from contaminated food.

We should first note that the pooled prevalence of intestinal carriage of drug-resistant Enterobacteriaceae in our review (14% for all Enterobacteriaceae and 18% for ESBL producing Enterobacteriaceae) has slightly increased from an earlier review (14% [95% CI 9-20%] for ESBL producing Enterobacteriaceae) published in 2016 (34). Karanika *et al*. conducted a systematic review and meta-analysis on papers published from 1978 to 2015 under search terms "ESBL" or "extended-spectrum beta-lactamase", and limited the studies conducted in OECD countries. Our literature search was not limited to ESBL producing bacteria nor OECD countries. Some studies reported carbapenemase-producing Enterobacteriaceae (CPE), and extended-spectrum cephalosporin (ESC) resistant *E. coli*. High variability in the prevalence among studies could be explained by infections from external sources such as the environment, contaminated food, and contaminated water, in addition to high variability in antimicrobial usage in different regions of the world.

The high variability could also be explained by the types of populations studied. In our study, the prevalence between travellers and general adult populations were significantly different (8% [95% CI 4-14%] and 37% [95% CI 30-43%]), respectively, suggesting different mechanisms for acquiring drug-resistant gut Enterobacteriaceae organisms. It is possible that travel includes distinct behavioral activities that affect exposure to potential risk factors for acquiring ARGs. This assumption led us to examine the pooled estimates of OR for each risk factor stratified by traveler vs general adult population.

We found four risk factors significantly associated with intestinal carriage of drug-resistant *E. coli.* Prior antimicrobial drug use within 12 months prior to stool culture, diarrhea symptoms, and travel to India were significant risk factors, which were also identified in previous reports (34, 35). When controlled by travel status, we found the same risk factors (antimicrobial use, diarrhea, diet and travel to India) significantly associated with fecal carriage of drug-resistant *E. coli* for travellers. Meanwhile, for the general adult population, travel to Southeast Asia was significantly associated with ARG carriage in addition to antimicrobial use and diarrhea. We should note that due to the limited number of studies, some risk factors commonly assessed for entire populations could not be assessed for stratified populations.

To the best of our knowledge, no previous review has found vegetarian diet to be significantly

associated with intestinal carriage of drug-resistant *E. coli*. Butcher *et al*. (2019) reported that unwashed vegetables could be a source for ESBL-producing extraintestinal pathogenic *E. coli* (80). Multiple reports suggest association between urinary pathogenic *E. coli* and fecal *E. coli* (81, 82), and fecal carriage of drug-resistant *E. coli*. Although we should note that our pooled ORs for drug-resistant *E. coli* intestinal carriage were not controlled for potential confounding factors other than travel status, our findings suggest that certain type of dietary practice could be a risk factor for acquiring drug-resistant *E. coli* by the gut microbiota.

In addition to these four significant risk factors, we identified ten other risk factors commonly assessed in 3 or more reviewed studies. These include gender, smoking, living with pet(s), education level, proton-pump inhibitor use, previous hospital admission, chronic disease, international travel, travel to Southeast Asia, and travel to Africa. None of these factors were significantly associated with risk of intestinal carriage of drug-resistant *E. coli*.

However, 50% or more of the studies reported significant associations for proton-pump inhibitor use, chronic disease, and travel to Southeast Asia. This suggests that these factors could serve as risks for drug-resistant *E. coli* colonization under certain situations. In fact, travel to Southeast Asia was a significant risk factor for general adult populations. Previous hospitalization and travel to Africa were also assessed in the review by Karanika *et al* (34). In agreement with our findings, previous hospitalization and travel to Africa were not significant risks. Stratification based on location of studies such as OECD countries to non-OECD countries and features of travel destination such as sanitation system and antibiotics usage in food production can alter the pooled ORs.

Multiple studies reported food as potential sources of *E. coli* infections (25, 26, 28, 53, 80). To the best of our knowledge, we found no other reviews that examined the effect of food on fecal carriage of drug-resistant *E. coli*. Being a vegetarian was significantly associated with the carriage of drug-resistant *E. coli* among overall population and travellers. Pooled estimate among general adult populations could not be obtained due to limited number of studies. Several recent studies have reported contamination of leafy green vegetables with saprophytic bacteria harboring ARGs that occur in human Gram-negative bacterial pathogens (53, 83, 84). Four studies reported the effect of street food, raw vegetables, and raw milk consumption (65, 72, 75, 76). However, these factors showed high variance in reported ORs among studies. This variance could be explained by differences in study region, target population, travel destination and sanitation conditions among studies. One study reported conflicting ORs for raw vegetable consumption between Southeast Asia (Brunei Darussalam, Cambodia, Indonesia, Lao People's Democratic Republic, Malaysia, Myanmar, Philippines, Singapore, Thailand, Timor-Leste, Viet Nam) and South Asia (Afghanistan, Bangladesh, Bhutan, India, Iran (Islamic Republic of), Maldives, Nepal, Pakistan, and Sri Lanka) (65). Geographic differences in food production methods and

antimicrobial drug usage could exist. Although further studies on vegetable consumption among general population are required, this observation suggests that dietary habit can affect fecal carriage of drug-resistant *E. coli*, which supports our hypothesis that, in addition to healthcare-associated acquisition and person-to-person transmission, ARGs may be acquired via contaminated food.

There are limitations associated with this systematic literature review. First, some studies investigated Enterobacteriaceae instead of *E. coli* alone. Still, the frequency of *E. coli* found among studies that examined Enterobacteriaceae was high (79-97%). One study that had low frequency (29%) of *E. coli* was not eligible for meta-analysis. Therefore, we can assume that risk factors identified in this review would apply to *E. coli*.

Also, we cannot determine whether the identified risk factors have causal effects on fecal carriage of drug-resistant *E. coli*. For example, an episode of diarrhea among participants could have prompted the use of antibiotics, which could have selected for drug-resistant *E. coli* in the host intestinal microbiota. Still, identification of factors significantly associated with the carriage of drug-resistant *E. coli* will be useful for identifying individuals with high risk and early focused interventions. Another limitation of our study is that there was no study from North America included in this review. Karanika *et al*. (2016) reported the same limitation (34). Since North America is a major food-exporting region in which antibiotics are heavily used in food animal husbandry and agriculture, if food is an important reservoir for drug-resistant bacteria that enter our intestines, more studies in this geographic region are needed. Also, although we did not observe publication bias for risk factors identified in this study, we found high heterogeneity among studies that reported the risk of chronic disease and travel related factors on intestinal carriage of drug-resistant bacteria. This high heterogeneity could be explained by differences in sampling methods, chronic diseases reported, travel destinations, and sanitation conditions examined in the studies. These differences could have affected the pooled OR estimates. Particularly, we should note that the chronic diseases three studies investigated were different among studies, and there was a high variation in disease incidence within the studies (65, 70, 78). Furthermore, there were three studies reporting association for PPI use as risk factors for fecal carriage of drug-resistant *E. coli* (70, 71, 79), and McNulty *et al*. (2018) stated in their limitation that they did not collect data on the use of PPI (76). Since PPI use is one of the indicators of chronic disease, larger studies related to PPI use and other chronic diseases may alter the result.

In this review, we found five significant risk factors associated with intestinal carriage of drug-resistant *E. coli*. Due to the high heterogeneity of the studies, other factors may indeed serve as risks under certain circumstances. Further studies, especially those that examine food and other environmental exposures will be essential for identifying public health interventions that can be devised to decrease human intestinal colonization with drug-resistant bacteria.

Tables

Table 1: Characteristics of studies included in review, 2014-2019

Author, year	Country	Study population	Study design	Study period	Sample size	Pathogen type
Arcilla 2017	Netherlands	Travellers	Prospective cohort study	2012 Nov - 2013 Nov	1847	ESBL-PE, CPE
Angelin 2015	Sweden	Travellers	Prospective study	2010 Apr - 2014 Jan	99	*E. coli*
Caudell 2018*	Tanzania	General adult	Prospective study	2012 Mar - 2015 Jul	226*	*E. coli*
Dohmen 2017	Netherlands	Employees in a pig slaughterhouse	Prospective study	2015 Jun	334	*E. coli*
Dohmen 2017*	Netherlands	Pig farmers, family members and employees	Longitudinal study	2011 Mar - 2011 Oct	146	ESBL-PE
Lubbert 2015	Germany	Travellers	Prospective cohort study	2013 May - 2014 Apr	191	ESBL-PE
McNulty 2018	England	General adult	Retrospective cohort study	2013 - 2014	2430	ESBL-PE
Miranda 2016	Germany	Travellers	Retrospective study	2013 Feb - 2014 Apr	211	ESBL-PE
Mo 2019	Singapore	General adult	Cross sectional study	2016 Jun - 2017 Apr	305	ESBL-PE
Reuland 2016	Netherlands	General adult	Case control study	2011 Jun - 2011 Nov	1695	ESBL-PE
Reuland 2015	Netherlands	General adult	Case control study	2011 Aug - 2011 Dec	550	pAmpC producing *E. coli*
Ruh 2019	Northern Cyprus	General adult	Retrospective cohort study	2017 Sep - 2017 Dec	500	Enterobacteriaceae
Sanneh 2018*	Gambia	Food handlers	Cross sectional study	2015 Jul - 2015 Sep	565	Enterobacteriaceae
Vading 2016	Sweden	Travellers	Prospective cohort study	2013 Apr - 2015 May	175	ESBL *E. coli*
Wielders 2017	Netherlands	General adult	Cross sectional study	2012 Nov	2432	ESBL-PE
* not included in meta-analysis						

Note: ** not included in meta-analysis. *indicates sample size was households (all others are individuals). ESBL-PE = Extended-spectrum beta-lactamase producing Enterobacteriaceae; CPE = Carbapenemase-producing Enterobacteriaceae

Table 2: Commonly assessed risk factors for intestinal carriage of drug-resistant *E. coli*, 2014-2019.

Risk Factor	Number of studies investigated*	Number of studies finding significant association*	Number of samples assessed	Number of samples with drug resistant bacteria	Pooled OR (95%CI)	χ^2 (P-value)	I^2
General factors							
Gender	11	1	9836	1428	1.16 (0.98-1.36)	8.77（0.46)	0
Diet restriction (vegetarian)	5	2	6802	989	1.60 (1.00-2.56)	3.22 (0.52)	0
Pet	4	1	5159	407	1.15 (0.33-4.06)	5.23（0.16）	43
Education level	4	0	5067	925	0.93 (0.74-1.17)	0.98 (0.81)	0
Smoking	4	1	4497	712	0.77 (0.18-3.25)	6.37 (0.04)	69
Clinical factors							
Antimicrobial use	13	6	10079	1407	1.84 (1.35-2.51)	18.28 (0.05)	45
Previous hospital admission	7	2	6108	465	1.63 (0.84-3.18)	7.83 (0.17)	36
Diarrhea	7	4	5144	1079	1.56 (1.09-2.25)	5.76 (0.33)	13
Proton-pump inhibitor use	3	2	4111	359	1.31 (0.11-15.5)	5.81 (0.05)	66
Chronic disease	3	2	2323	766	0.91 (0.13-6.53)	8.68 (0.01)	77
Travel related factors							
International travel	6	2	6460	520	1.13 (0.67-1.91)	10.73 (0.06)	53
Travel to Southeast Asia	8	4	6632	1289	1.78 (0.64-4.98)	50.28 (<0.01)	86
Travel to Africa	5	2	6692	1105	1.29 (0.52-3.21)	81.34 (<0.01)	94
Travel to India	4	4	2953	423	4.15 (2.54-6.78)	2.50 (0.48)	0

OR = Odds Ratio; CI = Confidence interval. Note: *indicates results from systematic review

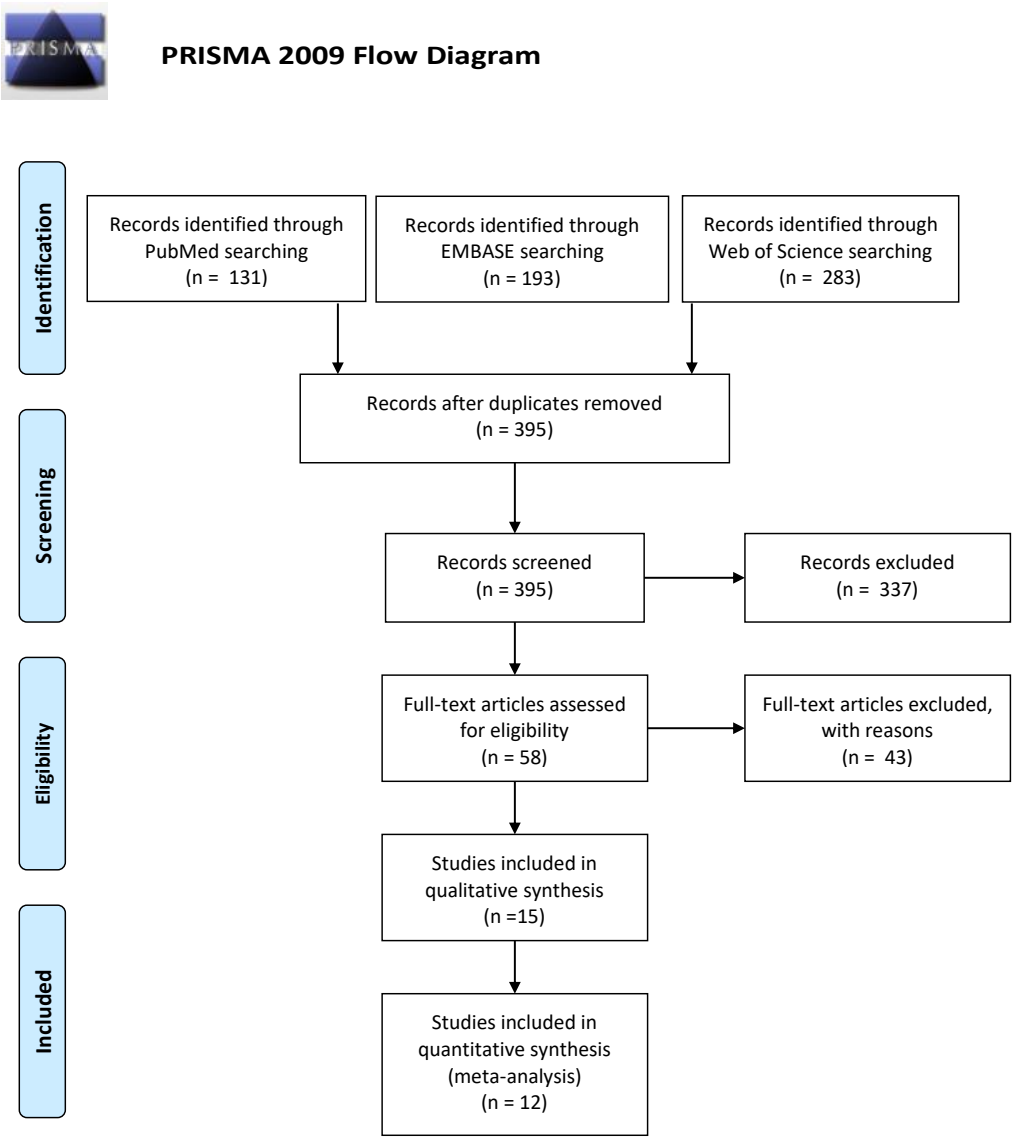
Table 3: Commonly assessed risk factors for intestinal carriage of drug-resistant *E. coli*, 2014-2019, stratified by travellers and general adults

	Travellers				General adults			
Risk Factor	Number of studies investigated *	Pooled OR (95%CI)	χ^2 (P-value)	I^2 (%)	Number of studies investigated *	Pooled OR (95%CI)	χ^2 (P-value)	I^2 (%)
General factors								
Gender	4	1.14 (0.85-1.51)	2.17 （0.54)	0	6	1.16 (0.90-1.50)	6.15 (0.29)	19
Diet restriction (vegetarian)	3	1.92 (1.13-3.26)	1.29 (0.52)	0	1	-	-	-
Pet	1	-	-	-	3	0.93 (0.70-1.24)	0.94 (0.63)	0
Education level	1	-	-	-	3	0.92 (0.63-1.35)	0.98 (0.81)	0
Clinical factors								
Antimicrobial use	4	2.81 (1.47-5.36)	4.07 (0.25)	26	7	1.51 (1.17-1.94)	5.54 (0.48)	0
Previous hospital admission	1	-	-	-	5	1.47 (0.79-2.76)	5.54 (0.24)	28
Diarrhea	4	1.65 (1.02-2.68)	5.16 (0.16)	42	3	1.53 (1.27-1.84)	0.43 (0.80)	0
Travel related factors								
International travel	0	-	-	-	6	1.13 (0.73-1.74)	10.73 (0.06)	53
Travel to Southeast Asia	5	1.93 (0.46-8.12)	41.24 (<0.01)	90	8	1.67 (1.02-2.73)	5.56 (0.06)	64
Travel to Africa	3	0.75 (0.29-1.96)	19.27 (<0.01)	90	2	-	-	-
Travel to India	3	3.80 (2.23-6.47)	1.62 (0.45)	0	1	-	-	-

OR = Odds Ratio; CI = Confidence interval. Note: *indicates results from systematic review

Figures

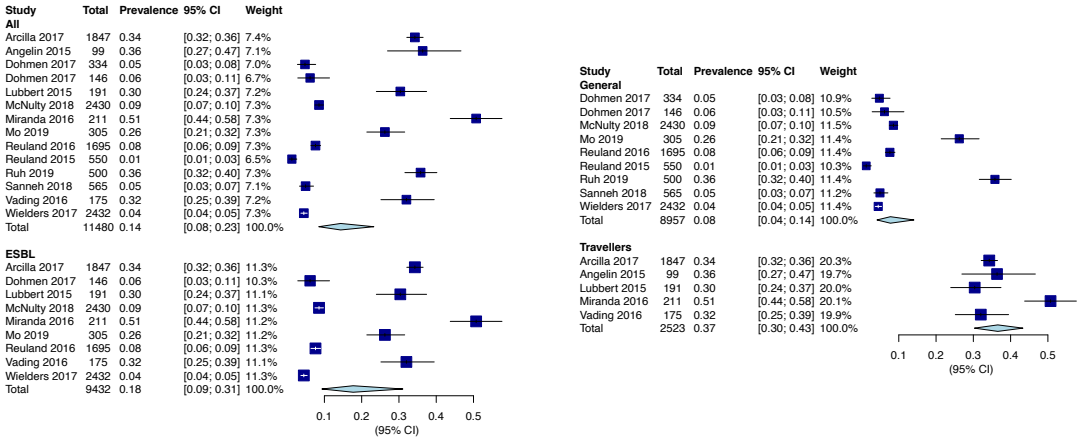
Figure 1: PRISMA Flow Diagram. Flow diagram of the systematic review process used to identify eligible studies



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

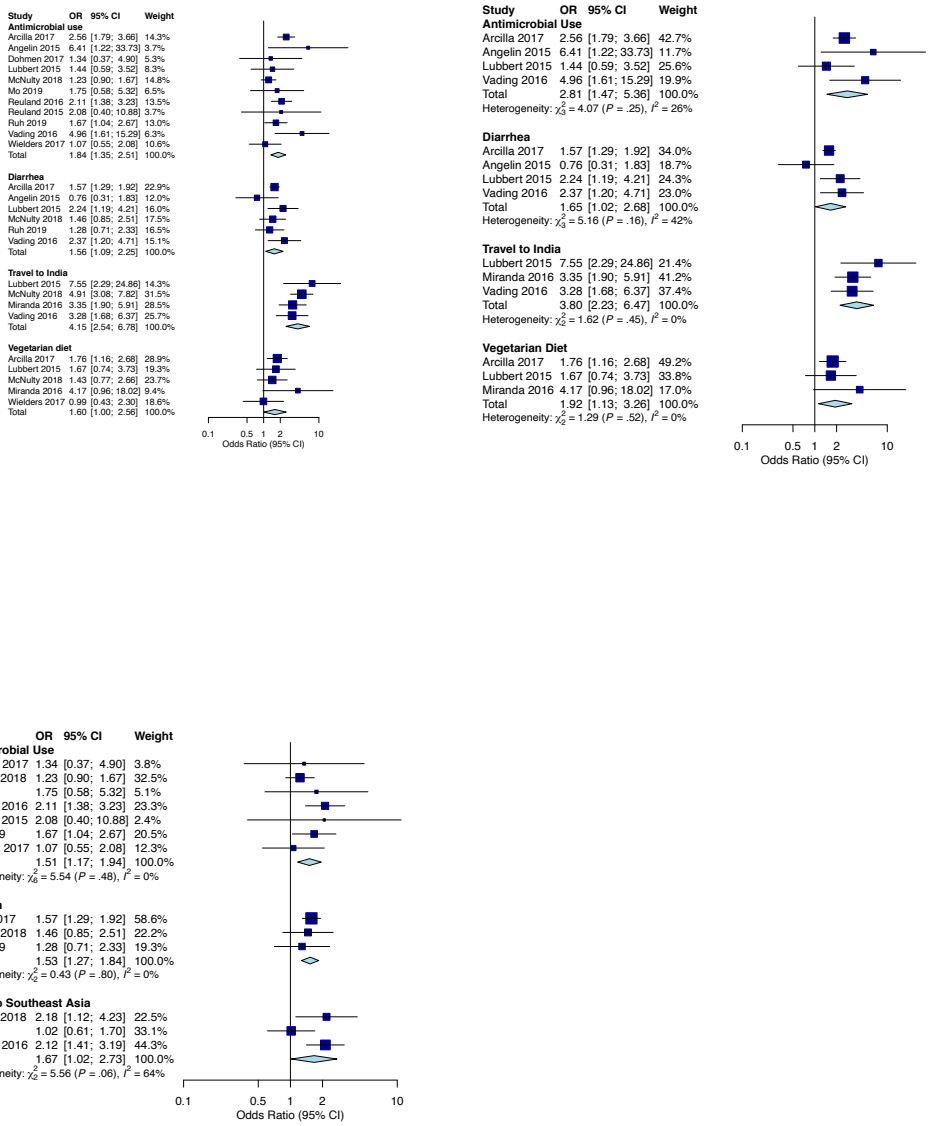
Figure 2: Forest plots for individuals and combined prevalence estimates of fecal carriage of drug-resistant bacteria.



a

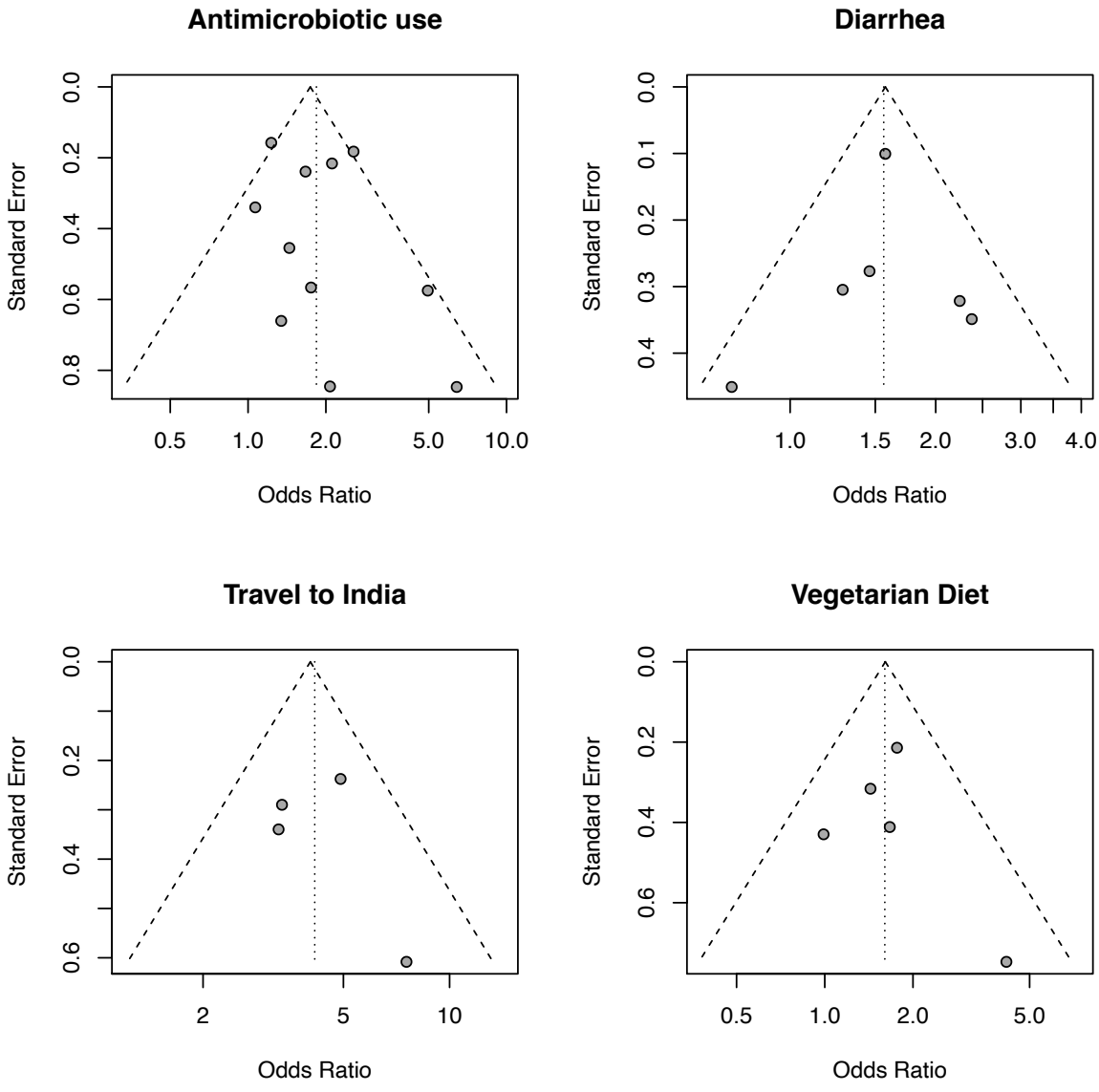
Prevalence of drug-resistant Enterobacteriaceae and prevalence of ESBL-producing Enterobacteriaceae; **b** Prevalence of drug-resistant Enterobacteriaceae among travellers and general populations

Figure 3: Forest plots for significant risk factors.



a Individuals and combined OR of fecal carriage of drug-resistant *E. coli* among entire population; **b** Individuals and combined OR of fecal carriage of drug-resistant *E. coli* among travellers; **c** Individuals and combined OR of fecal carriage of drug-resistant *E. coli* among general population. OR, odds ratio

Figure 4: Funnel plots



Funnel plots for studies reporting antimicrobial use, diarrhea, vegetarian diet, and travel to India as risk factors

Chapter 3: Risk factors for fecal carriage of multidrug-resistant *Escherichia coli* in a college community: a penalized regression model

A version of this chapter was published in the Journal of Global Antimicrobial Resistance in September 2021.
Reference: Hu, Y., Rubin, J., Mussio, K. & Riley, L. W. Risk factors for fecal carriage of multidrug-resistant *Escherichia coli* in a college community: a penalized regression model. Journal of Global Antimicrobial Resistance 26, pp166-173 (2021). https://doi.org/10.1016/j.jgar.2021.05.004.

Introduction

Antimicrobial drug resistance is one of the most pressing global public health concerns of our time, threatening the effective prevention and treatment of infectious diseases in every country (85, 86). In the United States, the U.S. Department of Health & Human Services has declared the necessity to take national action and to strengthen surveillance systems to combat the spread of antimicrobial resistance (87). In particular, several drug-resistant Gram-negative bacterial species belonging to Enterobacteriaceae have come to be designated by the Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO) to be "urgent-threat" or "priority 1" pathogens (51).

Among the Gram-negative bacterial pathogens, *Escherichia coli* (*E. coli*) is the most frequent cause of common extraintestinal infections, including urinary tract infections (UTI) and bloodstream infections (BSI) (17). They are referred to as extraintestinal pathogenic *E. coli* (ExPEC) (6). The prevalence and incidence of infections caused by drug-resistant ExPEC have been rapidly increasing worldwide (22, 23).

Major sources of drug-resistant Gram-negative bacteria include the environment such as contaminated water (24), food including meat (25, 26) and vegetables (28, 53), and healthcare settings (29). Additionally, intestinal commensal drug-resistant bacteria have been reported as an important reservoir of antimicrobial drug resistance genes (ARGs) (54, 55). *E. coli* is a member of the commensal flora of human and other warm-blooded animal intestinal tracts. As such, they can acquire ARGs by horizontal gene transfer (39) from drug-resistant *E. coli* strains and other Gram-negative bacteria that enter the intestinal tract via exposures to contaminated food, water, and other external sources. *E. coli* can be transmitted through contaminated water or food, or through contact with people and other animals (21). Thus, risk factors for fecal carriage of drug-resistant commensal *E. coli* and ARGs could include exposures to environmental sources of drug-resistant bacteria, in addition to traditional risks such as prior use of antibiotics (33, 88) or

healthcare-associated infections (36, 37).

In our previous study, we found 81% of healthy volunteers at a college community to harbor intestinal drug-resistant commensal Gram-negative bacteria (89). They included 12 species of Gram-negative bacteria that harbored a variety of ARGs and integrons that are frequently found among extraintestinal pathogenic *E. coli* (ExPEC) (89). Risk factors for the carriage of these ARGs, however, were not determined.

We first conducted a systematic review and meta-analysis on risk factors of intestinal carriage of drug-resistant *E. coli* (90), but found no studies from North America that met the criteria to be included in the review, even though North America is a major food-exporting region in which antibiotics are heavily used in food animal husbandry and agriculture. Also, we could not find studies that particularly focused on dietary behavior as a risk factor for intestinal carriage of drug-resistant *E. coli* or ARGs. Two studies from The Netherlands reported a significant association between food and carriage of drug-resistant bacteria. One study reported that raw milk consumption increases the risk of intestinal colonization of drug-resistant bacteria (72), and another reported food consumption from street vendors as a travel-associated risk factor (65).

Here, we designed and implemented a cross-sectional study to identify dietary risk factors associated with fecal carriage of multidrug-resistant commensal *E. coli* among healthy adult population in a US college community. We focused on the carriage of multidrug-resistant *E. coli* in particular because of its clinical relevance to diseases such as UTI and BSI.

Methods

Sample collection, antimicrobial susceptibility screening, and questionnaires

As previously described by Rubin et al. (89), we prospectively collected and cultured fecal swab samples from 113 healthy volunteers at a university campus in northern California between June and October 2018. Volunteers were recruited from students and staff frequenting a campus café and people working in a campus health science research building. After the objectives of the survey were explained, volunteers who consented to the study completed and submitted the survey instrument. Eligible participants included both men and women between 18 and 65 years of age, with no medical history of urinary tract corrective surgery or abnormality, or bladder catheterizing or hospitalization within the 6 months prior to sample collection. At recruitment, participants were provided a collection kit containing a Blair transport media rectal swab (Becton Dickinson BBL), two bio-safety bags, and detailed collection instructions. Each kit also included a questionnaire regarding antibiotic use, history of UTI, as well as diet and lifestyle characteristics. Participants were instructed to send the swab and the completed questionnaire back to the laboratory via USPS mail immediately after collection. Once delivered, the study coordinator analyzed the fecal swab samples within 48 hours. Detailed microbiologic procedures, including antimicrobial drug susceptibility testing of the *E. coli* isolates, are described in the report by Rubin et al (89). We tested the *E. coli* for drug susceptibility to ampicillin (Amp), trimethoprim-sulfamethoxazole (SXT), gentamicin (CN), and colistin (COL), each representing a different class of antimicrobial agent.

Survey

The survey questions focused on dietary and behavioral habits that might be associated with the intestinal acquisition of ARG. We asked participants to recall their habits within the 1-year period before completion of the survey questionnaire. The survey instrument was devised partly based on the standard questionnaire used by the Centers for Disease Control and Prevention, Foodborne Disease Outbreak and Surveillance Unit, National Hypothesis Generating Questionnaire (https://www.cdc.gov/foodsafety/outbreaks/surveillance-reporting/investigation-toolkit.html) and was partly based on the instrument used by Manges et al. in a 2007 study investigating the links between retail meat and risk of UTI (91). Participants were asked to respond according to the following scale: never, past day (24 hours), past week, past month, past year, or more than a year ago. Behavior within the past month was defined as frequent. The following information was requested of the participants: biological sex, academic status, frequently used supermarkets, meat consumption, dietary restrictions, dairy and egg consumption, raw meat consumption, raw vegetable consumption, preference of organically produced food, number of live-in housemates, location of meal preparation, possession of companion animals, number of sexual partners, history

of antibiotic use, history of hospitalization, history of UTI, placement, and type of an intrauterine device, and location and duration of travel outside of the United States.

Dataset construction

To identify the risk factors associated with intestinal carriage of antimicrobial resistance, we compiled a set of 66 variables from our survey that detailed their diet habit, lifestyle, and past antibiotic use. All questions were dichotomized based on frequent ("Yes") or modest ("No") responses as described above. We excluded variables for which (1) less than 5% or more than 95% of the participants replied "Yes" and (2) more than 10% of the participants answered "Do not know" or NA. We merged the results from antimicrobial resistance testing with the survey results. Relationships between each set of variables were assessed with correlation coefficients. In this study, we defined isolates with resistance to two or more antimicrobial drug classes as multidrug-resistant and those with resistance to only one drug as drug-resistant.

Statistical analysis

All statistical analyses were conducted with R version 3.5.1 (63). First, univariate and multivariate logistic regression analyses were performed. In the multivariate analysis, a model with all potential risk factors was constructed. Then we used R package 'stats' (version 3.5.1) 'step' function with a backward elimination method to determine the best-fitting model. The best-fitting model was determined with Akaike information criteria (AIC) (92). All tests performed were two-sided, and a P-value <0.05 was considered significant.

Penalized regression with the least absolute shrinkage and selection operator (LASSO) models (93) were used to identify variables associated with the carriage of multidrug-resistant *E. coli*. Correlation analysis was conducted to test the reliability of LASSO because correlations among potential risk factors would influence the prediction by LASSO. Association with drug resistance was not tested with LASSO regression due to the high imbalance between the prevalence of drug-resistant *E. coli* (84%) and drug-susceptible *E. coli* (16%). The LASSO regression functioned as a variable selection process, which reduced the variables to a subset of variables that were consistently related to multidrug resistance. We conducted 10-fold cross-validation to select the best fitting model with the minimum mean squared error (MSE). Standard errors were not assessed for LASSO regression model because of its strong bias arising from penalized estimation method. All LASSO regression was performed with R package 'glmnet' (version 2.0.18).

The model performance was evaluated by receiver operating characteristic (ROC) curves and the area under the curve (AUC) was used to classify the participants with and without multidrug-resistant *E. coli*. Hosmer-Lemeshow fit test was used to assess the agreement between observed and model-predicted proportions of carriage of multidrug-resistant *E. coli* (94). The difference of

AUCs was tested by a nonparametric approach developed by DeLong et al (95).

Results

Background information

Between June and October 2018, 113 fecal swab samples were collected and cultured, and Gram-negative bacteria were isolated from each culture. Of 113 volunteers, 103 returned both stool sample and the questionnaire. Of 103 stool samples, 93 yielded *E. coli* colonies on MacConkey plates. Ten rectal swabs showed no *E. coli* growth, possibly due to improper sample collection procedure or bacterial death during transport. Of 93, 81 corresponding survey questionnaires had complete responses without unanswered questions (Figure 1).

The participants were adults between 18 to 65 years of age (Table 1). Seventy-eight (84%) of 93 participants had *E. coli* resistant to either ampicillin, trimethoprim-sulfamethoxazole, gentamicin, or colistin. Of these, 48 (52%) were multidrug resistant, defined here as resistance to two or more classes of antimicrobial agents, and 30 (32%) were resistant to only one drug class. Fifteen (16%) were susceptible to all the tested drugs.

We compiled the responses to our questionnaire into a dataset with 66 variables and 103 observations (Supplementary Table 1). Variables with (1) more than 10 "Do not know" or (2) 5% or less "Yes" or "No" were removed from the data to be analyzed. Twenty-six (39%) of 66 variables were eligible for analysis. The prevalence of drug-resistant and multidrug-resistant *E. coli* isolates is shown in Table 1, and those of different class of antimicrobial agent is shown in Supplementary Table 2 among each category of the survey questions.

Relationships between each of the predictors under consideration are presented in Figure 2 as a correlation coefficient matrix. Females were more likely to report past urinary tract infection (UTI). Those reporting previous antibiotics use had a strong correlation with previous UTI. We observed a correlation between employment status and travel status. Also, dietary behaviors were correlated among each factor. There were positive correlations among red meat consumption, poultry consumption, but negative correlations among those reported to have dietary restrictions such as being vegan or vegetarian.

Univariate Analysis

The risk factors associated with the fecal carriage of any drug-resistant and multidrug-resistant *E. coli* based on univariate analyses are shown in Table 1. Any drug resistance was compared to those carrying drug-susceptible *E. coli*, while multidrug-resistance was compared to drug-susceptible and any drug-resistant *E. coli*. For the carriage of drug-resistant *E. coli*, frequent usage of Supermarket A and previous antibiotic use had a significant association (OR=3.4[95% confidence interval (CI95%)1.1-12.4], 4.6 [1.1-31], respectively). All participants who reported to have had past UTI carried drug-resistant *E. coli*. Female gender, undergraduate status, and red meat consumption were significantly associated with carriage of multidrug-resistant *E. coli* (OR = 2.8

[1.1-7.2], 3.9 [1.3-13.2], 2.6 [1.0-6.8], respectively).

Multivariate Analysis

The risk factors associated with the fecal carriage of drug and multidrug-resistant *E. coli* were assessed based on a multivariate logistic model (Table 2, Supplementary Table 3). The best-fitting multivariate logistic model was determined with a backward elimination method based on AIC.

Fish consumption was negatively associated with the carriage of drug-resistant *E. coli*, (OR = 0.17 [0.03-0.78]) whereas frequent usage of Supermarket A was a risk factor for drug-resistant *E. coli* colonization (OR = 4.5 [1.1-23]). Female gender (OR=2.3 [0.76-7.4]), organic food consumption (OR = 5.7 [0.67-150]), frequent usage of Supermarket B (OR = 3.5 [0.76-22]), and previous antibiotics use (OR = 4.3 [0.89-32]) included in the model showed a positive association with the carriage of drug-resistant *E. coli*, but the association was not significant when adjusted for other variables included in the model.

The variables significantly associated with the carriage of multidrug-resistant *E. coli* include female gender (OR=6.1 [1.9-22]), being an undergraduate student (OR=5.5 [1.5-25]), and frequent red meat consumption (OR=6.1 [1.8-24]). Frequent fish consumption was negatively associated with multidrug-resistant *E. coli* carriage (OR = 0.27 [0.08-0.85]). Frequent usage of Supermarket B (OR = 2.6 [0.87-8.3]) and previous travel (OR=2.3 [0.76-7.4]) included in the model showed a positive association with the carriage of multidrug-resistant *E. coli*, but the association was not significant when adjusted for other variables included in the model.

Penalized regression with LASSO

We performed penalized regression with LASSO for parameter selection. Ten-fold cross-validation was conducted to select optimal estimators. At sigma = 0.05823413, the mean squared error (MSE) showed the minimum and therefore we selected coefficients at sigma = 0.05823413 as optimum estimators (Figure 3). Estimators are shown in Table 2. Parameters included were female gender (OR=1.75), undergraduate student (OR = 1.96), frequent red meat consumption (OR = 1.82), frequent fish consumption (OR = 0.81), frequent usage of Supermarket B (OR = 1.07), and previous travel to Europe (OR = 1.12). Of these, all parameters other than previous travel to Europe were consistent with the results obtained from the multivariate model selection based on AIC.

Model comparison

The model performance was evaluated by receiver operating characteristic (ROC) curves (Figure 4). The area under the curves (AUC) of LASSO model and backward elimination method model

were 0.79 and 0.82, respectively (p=0.41). There was no difference in discrimination accuracy between the LASSO model and the model with a backward elimination method.

Discussions

This study identified potential risk factors associated with fecal carriage of multidrug-resistant *E. coli* among healthy adults at a northern California college community. As far as we know, this is the first study to find associations of dietary habits with carriage of multidrug-resistant *E. coli* in a healthy United States adult population. Previously, our fecal sample collection and surveillance conducted at the same college community from June to October 2018 identified the prevalence of drug-resistant Gram-negative bacteria among healthy volunteers and ARGs carried by these fecal bacteria (89). We reported the high prevalence of clinically common ARGs and integrons harbored by these Gram-negative bacteria (89). In the present study, we aimed to identify risk factors for the carriage of commensal drug-resistant *E. coli*.

Of all the submitted 103 fecal samples, 93 (90%) yielded *E. coli* on MacConkey plates. The absence of *E. coli* in the other ten samples could be due to the mishandling of the rectal swabs by volunteers or bacterial death during sample transport. Overall, the prevalence of drug-resistant *E. coli* was very high in our population, with 78 (84%) of 93 fecal samples showing resistance to at least one of the antimicrobial drugs tested, which included ampicillin, gentamicin, trimethoprim-sulfamethoxazole, and colistin. Of these, 48 (52%) showed resistance to at least two of these drugs. The prevalence of both drug-resistant and multidrug-resistant bacteria was higher than the pooled prevalence (14%) of a previous study (90) and even higher than the study with the highest prevalence (51%) that targeted a population of international travelers conducted in Germany (96). Unadjusted ORs revealed that while carriage of drug-resistant *E. coli* was positively associated with frequent usage of supermarket A and previous antibiotics use, carriage of multidrug-resistant *E. coli* was positively associated with female gender and frequent red meat consumption. We solicited data on frequented supermarkets in our survey instrument because of our previous study demonstrating the presence of drug-resistant *E. coli* in meat and produce products obtained from different retail markets in northern California (97, 98). These markets have different distributors and source farms where *E. coli* contamination could potentially occur. Other than the frequent usage of supermarket A, these findings are consistent with a previous report showing antibiotics use and female gender to be associated with the fecal carriage of drug-resistant bacteria (90) and multiple reports showing the isolation of multidrug-resistant *E. coli* from retail meat and produce (91, 97, 99).

We used multiple methods to identify risks. Since our sample size was small and was subject to overfitting under the multivariate logistic regression model, we used a penalized regression method LASSO as well to select features by shrinking less important coefficients to zero and to avoid overfitting (93). Still, features selected by LASSO are highly biased and unstable. Therefore, we implemented two methods to confirm the consistency of the outcome of model selection. Model selection for estimating the risk of multidrug-resistant *E. coli* carriage showed consistent

results with univariate analysis in both multivariate logistic regression backward elimination and LASSO regression methods. The backward elimination method with AIC and LASSO method both identified female gender, being an undergraduate student, frequent red meat consumption, and frequent usage of supermarket B as positively associated and frequent fish consumption as negatively associated with the carriage of multidrug-resistant *E. coli*. Factors related to previous travel were included in both models but were not significant nor consistent. Red meat consumption had an almost 6-fold increase and fish consumption had an almost 4-fold decrease in risk of the carriage of multidrug-resistant *E. coli,* when adjusted by gender, employment status, frequently used supermarket, and travel status. A similar trend is observed for the LASSO model, showing an 80% increase and 20% decrease for red meat and fish consumption, respectively, adjusted by gender, employment status, frequently used supermarket, and travel status.

Red meat has been recognized to contain *E. coli* strains harboring drug-resistant genes and therefore it was not surprising that frequent consumption of red meat was strongly associated with the carriage of multidrug-resistant bacteria (26, 99). Fish has also been reported to be colonized with multidrug-resistant bacteria and multiple drug resistance genes (100, 101), and yet we observed a negative association with both drug-resistant and multidrug-resistant *E. coli* carriage in multivariate logistic regression models. Another report found fish to protect against drug-resistant bacterial carriage(102) and other reports found the association of fish consumption with reduced relative abundance of Gram negative bacteria in gut microbiota (103, 104). From this study, we cannot draw any causal relationship between frequent fish consumption and low carriage of drug-resistant bacteria because of the potential unmeasured confounders related to frequent fish consumption and the carriage of drug-resistant bacteria.

Model performance of multivariate logistic regression and LASSO regression was compared with ROC. For our study, these two models' performance was equally moderate. One of the limitations of our study is the small sample size and the high imbalance between the prevalence of carriage of single-drug resistant bacteria and that of drug-susceptible bacteria. Although this strong imbalance itself suggests the serious public health impact of antimicrobial resistance in the study college community, we could not assess the risk factor for the carriage of any drug-resistant bacteria with the LASSO method. Still, we were able to determine risk factors associated with the carriage of multidrug-resistant bacteria based on the two regression methods.

A limitation of our study is that we relied on self-reports by the participants for an assessment of their dietary behavior. Our outcome of interest, the carriage of drug-resistant bacteria, was systematically tested in the laboratory. Still, our study was a cross-sectional study that simultaneously accessed the outcome and the exposures related to food. Therefore, we can assume overreporting or underreporting was non-differential and any measures of association would bias toward the null if any.

In conclusion, among healthy adults in a college community in Northern California, female gender, being an undergraduate student, and frequent consumption of red meat was significantly associated with increased risk of being colonized with multidrug-resistant *E. coli*, while frequent consumption of fish was negatively associated. Further studies with a larger population and at other locations will be essential for establishing the generalizability of our findings and to devise public health interventions that can decrease the colonization by drug-resistant bacteria.

Tables

Table 1: Univariate analysis of potential risk factors for faecal carriage of drug-resistant (DR) and multidrug-resistant (MDR) *Escherichia coli*

		DR (resistant to at least 1 antimicrobial drug)		MDR (resistant to at least 2 antimicrobial drug)	
	No. of population (%)	No. of samples with DR *E. coli* (%)	OR [CI95%]	No. of samples with MDR *E. coli* (%)	OR [CI95%]
Gender (Female)	49 (61)	43 (88)	2.39 [0.75-8.05]	29 (59)	2.77 [1.12-7.17]
Employment					
Undergrad student	19 (24)	18 (95)	4.78 [0.85-89.8]	14 (74)	3.88 [1.31-13.2]
Graduate student	37 (46)	27 (73)	0.27 [0.07-0.90]	15 (41)	0.52 [0.21-1.25]
Staff	23 (28)	20 (87)	1.56 [0.43-7.44]	10 (44)	0.72 [0.27-1.89]
Dietary habit					
Poultry	64 (79)	51 (80)	0.25 [0.01-1.38]	34 (53)	2.08 [0.70-6.67]
Red meat	52 (64)	42 (81)	0.67 [0.17-2.25]	30 (58)	2.59 [1.03-6.85]
Fish	31 (38)	23 (74)	0.39 [0.12-1.26]	12 (39)	0.50 [0.20-1.23]
Raw meat/fish	43 (53)	35 (81)	0.82 [0.25-2.61]	22 (51)	1.16 [0.49-2.81]
Raw vegetable	79 (98)	66 (84)	5.01 [0.19-134]	39 (49)	0.98 [0.04-25.2]
Diet restriction	19 (24)	18 (95)	4.78 [0.85-89.8]	8 (42)	0.68 [0.24-1.91]
Organic food	17 (21)	16 (95)	4.08 [0.72-76.9]	9 (53)	1.20 [0.41-3.57]
Food made at home	72 (89)	60 (83)	1.43 [0.20-6.83]	35 (49)	0.76 [0.18-3.08]
Food from outside home	30 (37)	25 (83)	1.07 [0.33-3.82]	16 (53)	1.29 [0.52-3.20]
Supermarket					
A	49 (61)	44 (90)	3.44 [1.06-12.4]	25 (51)	1.18 [0.48-2.90]
B	37 (46)	31 (84)	1.15 [0.36-3.83]	21 (57)	1.73 [0.72-4.23]
C	31 (38)	26 (84)	1.14 [0.35-4.06]	13 (42)	0.62 [0.25-1.51]
other grocery shop	31 (38)	26 (84)	1.14 [0.35-4.06]	13 (42)	0.62 [0.25-1.51]
Pet	22 (27)	17 (77)	0.61 [0.18-2.22]	10 (46)	0.81 [0.30-2.15]
Travel					
travel within one year	51 (63)	42 (82)	0.93 [0.26-3.02]	28 (55)	1.83 [0.74-4.65]
travel longer than 1 month	27 (33)	21 (78)	0.61 [0.18-2.22]	15 (56)	1.45 [0.57-3.72]
Southeast Asia	4 (5)	3 (75)	0.61 [0.07-12.9]	2 (50)	1.03 [0.12-8.91]
East Asia	10 (12)	7 (70)	0.43 [0.10-2.21]	5 (50)	1.03 [0.27-4.00]
Central and South America	21 (26)	18 (86)	1.35 [0.37-6.45]	10 (48)	0.91 [0.33-2.47]
Europe	27 (33)	23 (85)	1.31 [0.39-5.18]	16 (59)	1.82 [0.72-4.73]
Medical history					
previous antibiotics use	31 (38)	29 (94)	4.58 [1.13-30.9]	18 (58)	1.76 [0.72-4.43]
past UTI	12 (15)	12 (100)	NA	6 (50)	1.03 [0.30-3.60]
Others					
Living with roommate	71 (88)	59 (83)	1.23 [0.17-5.70]	36 (51)	1.54 [0.41-6.48]

OR, odds ratio; CI, confidence interval; UTI, urinary tract infection; NA, not applicable.

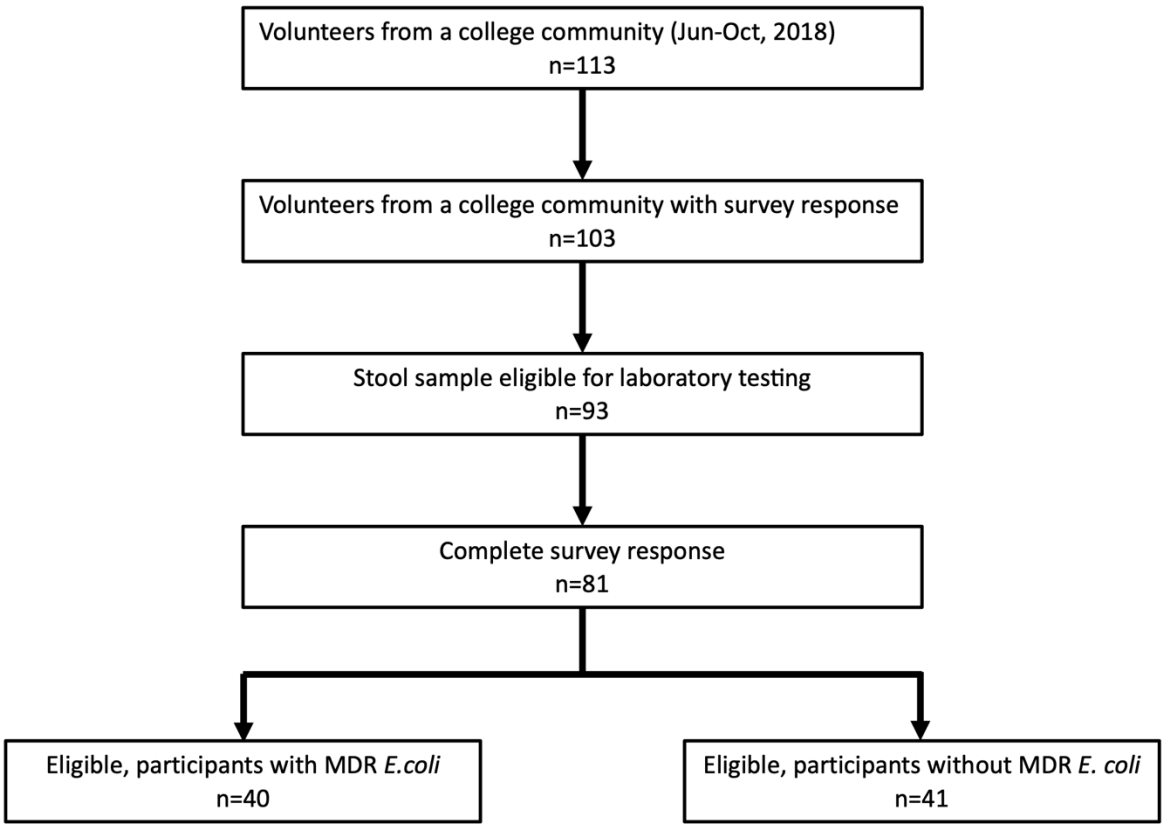
Table 2: Result of multivariate logistic regression with backward elimination model and LASSO model

	Multivariate logistic regression			LASSO regression
	OR	95%CI	P-value	OR
Female gender	6.06	1.92-22.5	0.004	1.75
Undergraduate	5.46	1.47-25.1	0.017	1.96
Red meat	6.13	1.83-24.2	0.005	1.82
Fish	0.27	0.08-0.85	0.032	0.81
Supermarket B	2.56	0.87-8.28	0.098	1.07
Travel within 1 year	2.29	0.87-8.28	0.14	-
Travel to Europe	-	-	-	1.12
AUC	0.82			0.79

LASSO, least absolute shrinkage and selection operator; OR, odds ratio; CI, confidence interval; AUC, area under the curve.

Figures

Figure 1: Flowchart of study samples



MDR, multidrug resistance

Figure 2: Correlation matrix for potential risk factors.

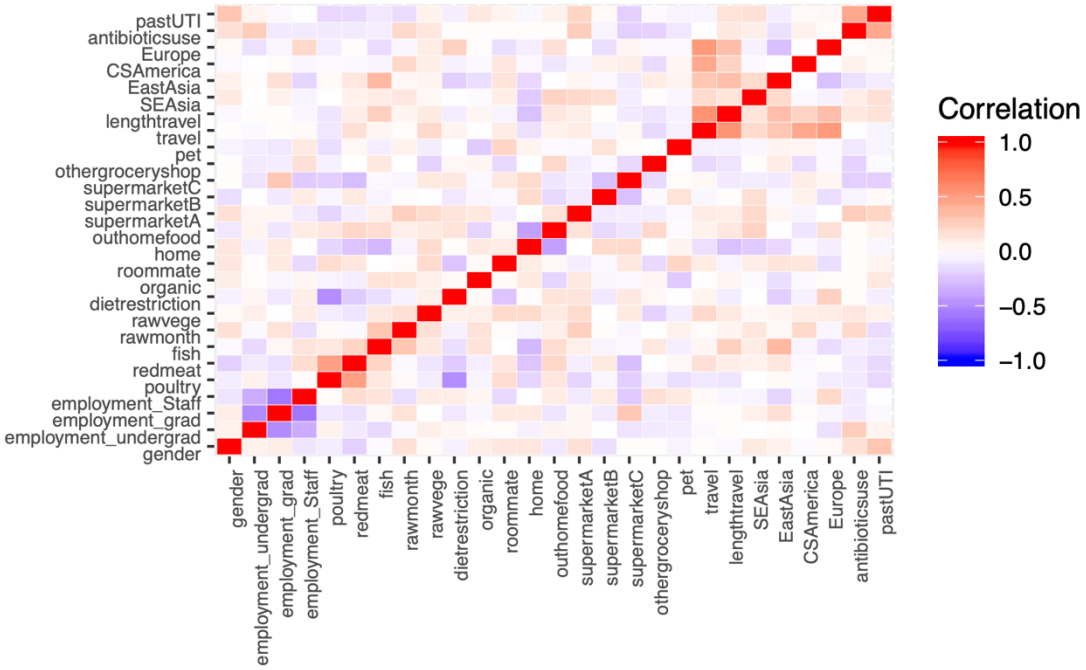
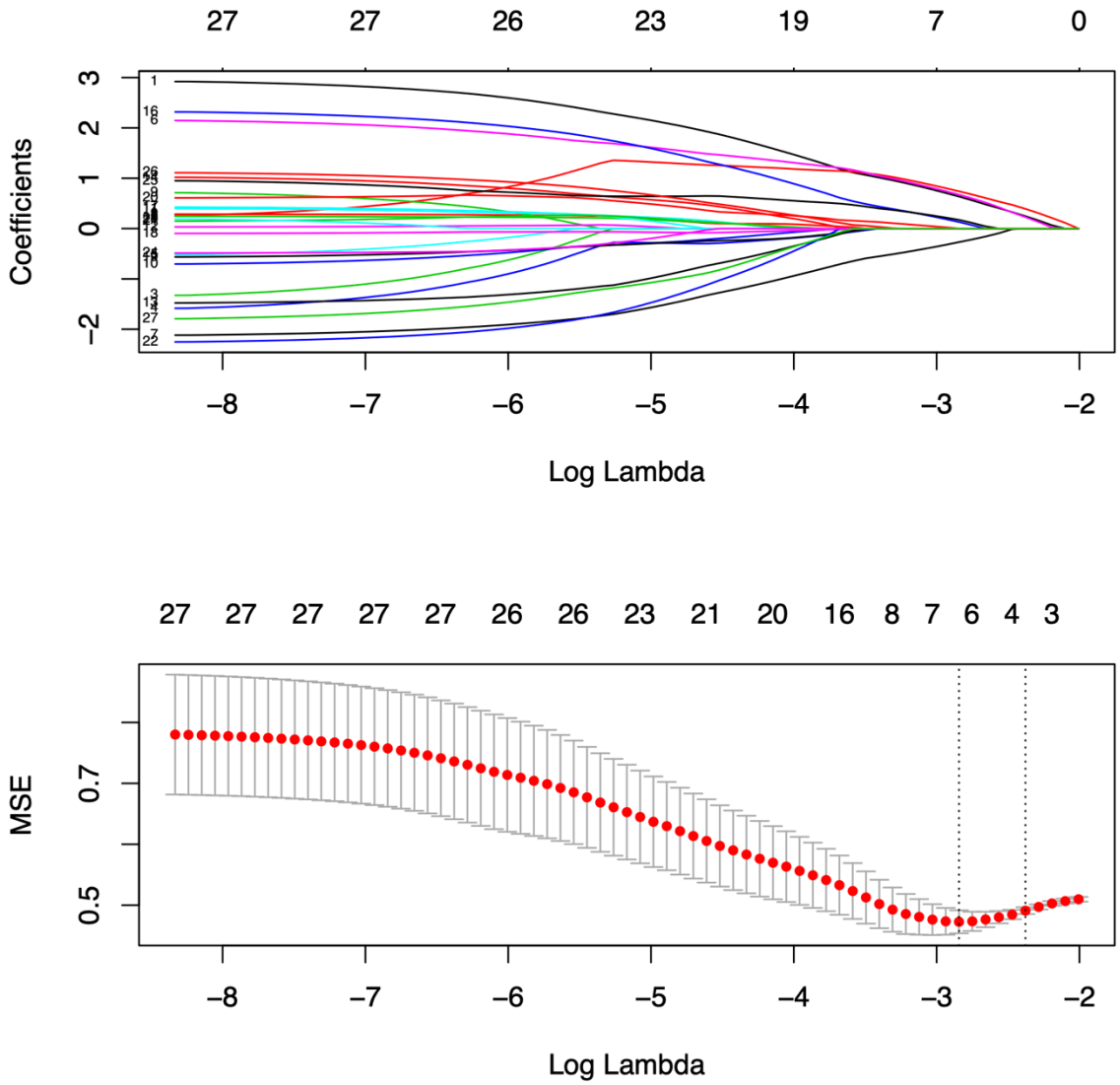
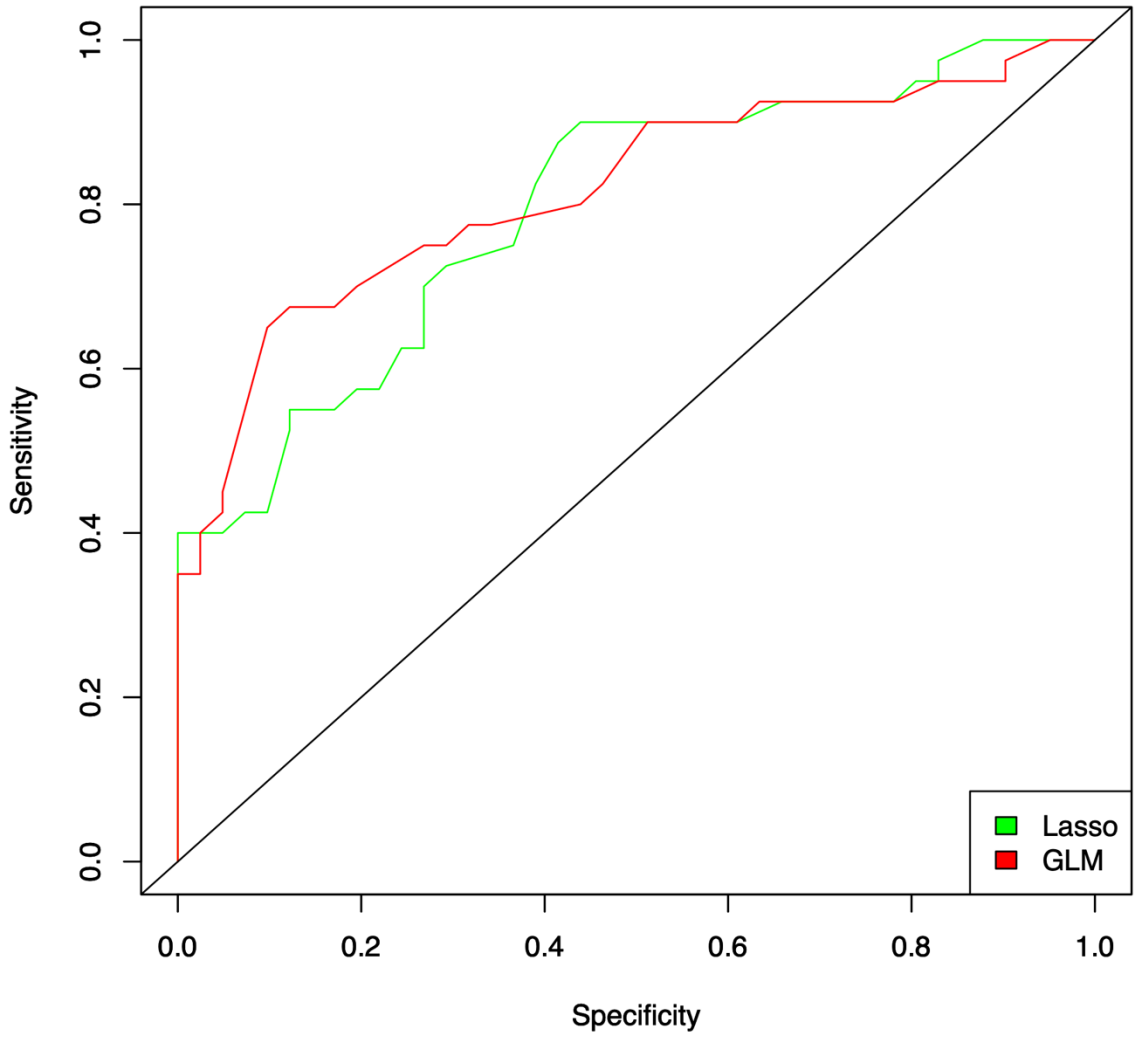


Figure 3: Plots for LASSO regression coefficients and cross-validation over different values of the penalty parameter.



MSE, mean squared error

Figure 4: Receiver operating characteristic (ROC) curve



Discrimination performance of LASSO model (green) for faecal carriage of multidrug-resistant Escherichia coli and comparison with generalised linear model (GLM) backward elimination method model (red) by ROC analyses.

Chapter 4: Genetic features of persistently antimicrobial drug-susceptible extraintestinal pathogenic *Escherichia coli* pandemic sequence type 95

Reference: Allegretti, Y.H., Yamaji, R., Adams-Sapper, S. & Riley, L. W. Genetic features of persistently antimicrobial drug-susceptible extraintestinal pathogenic *Escherichia coli* pandemic sequence type 95.
https://doi.org/10.1101/2021.10.28.466352

Introduction

Antimicrobial drug resistance is one of the most serious global public health concerns of our time, threatening the effective prevention and treatment of infectious diseases (85, 86). In particular, several drug-resistant Gram-negative bacterial species belonging to Enterobacteriaceae have come to be designated by the Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO) to be "urgent-threat" or "priority 1, critical" pathogens in need of new treatment and research (51, 85). Among Enterobacteriaceae, *Escherichia coli* (*E. coli*) is the most frequent cause of common extraintestinal infections such as urinary tract infections (UTI) and bloodstream infections (BSI) (17). They are referred to as extraintestinal pathogenic *E. coli* (ExPEC) (6). The prevalence and incidence of infections caused by ExPEC strains resistant to a wide spectrum of antimicrobial agents, including extended-spectrum beta-lactam (ESBL) drugs, fluoroquinolones, trimethoprim-sulfamethoxazole, and aminoglycosides have been rapidly increasing worldwide (9, 22, 23, 90, 105–107).

Epidemiological studies have shown that large proportions of community and healthcare-associated UTI and BSI reported globally are caused by ExPEC strains belonging to a limited set of lineages as defined by multilocus sequence typing (MLST) (14, 108). These include sequence types 131 (ST131), ST95, ST69, ST73, and ST393, which are responsible for nearly half of all *E. coli* UTIs or BSIs reported from many regions of the world (4). Most of these lineages except for ST95 have a high frequency of drug resistance. In particular, a large proportion of ST131, the most commonly reported ExPEC lineage worldwide, is multidrug-resistant (109–111).

Interestingly, despite its high prevalence, several recent studies have reported that ST95 strains remain susceptible to most of the antimicrobial drugs used to treat UTI and BSI (5, 108, 112, 113). We previously reported that the carriage of a UTI89-like plasmid called pUTI89* and *fimH* type is associated with the pan-susceptibility phenotype of ST95 (5, 113). However, the particular genomic features associated with this persistently drug-susceptible phenotype remained unclear. Here, we compared whole-genome sequences (WGS) of susceptible ST95 strains to those of ST95 strains resistant to at least one drug to identify genetic features

associated with persistent drug susceptibility. We hypothesize that persistently susceptible ST95 strains restrict the acquisition or maintenance of mobile DNA elements that carry drug-resistance genes, but they nevertheless may have a compensatory fitness advantage that determines their successful global dissemination.

Methods

Sample collection in Northern California

As previously described by Yamaji et al. (108) and Adams-Sapper et al. (113), we consecutively collected urine samples from patients with symptoms of UTI attending our university outpatient health service between September 2016 and May 2017, and all BSI isolates from inpatients admitted to San Francisco General Hospital (SFGH) between July 2007 and September 2010. All *E. coli* isolates were genotyped by multilocus sequence typing (MLST) based on the seven-gene scheme described at the PubMLST website (https://pubmlst.org/bigsdb?db⁄⁄pubmlst_mlst_seqdef) (114). All *E. coli* subtyped as ST95 were selected for WGS analysis. We analyzed 36 ST95 isolates from patients with UTI and 44 ST95 isolates from patients with BSI.

Short read genome sequencing

Genomic DNA was extracted from all strains with the Qiagen blood and tissue DNeasy kit. Library preparation for the MiSeq platform following a standard protocol for Illumina-compatible libraries was conducted at Functional Genomics Laboratory in Berkeley. Final libraries were sequenced via a 300-bp paired-end run on a MiSeq instrument with V3 chemistry and standard Illumina analysis software at The California Institute for Quantitative Biosciences at UC Berkeley (QB3-Berkeley).

Genome assembly

MiSeq reads were screened and trimmed based on length and quality with BBDUK version 1.0 under the default setting (http://jgi.doe.gov/data-and-tools/bb-tools/). The trimming process also removed residual adapter sequences. Quality control of individual FASTQ files was conducted with FastQC (115) and all of the newly sequenced WGSdata passed the quality check. *De novo* assembly of trimmed paired reads for the 80 libraries was performed with Unicycler version 0.4.8 under the default setting (116). The number of reads used in each assembly was sufficient to give a minimum of 25-fold coverage, averaged across all contigs. A maximum of 45-fold coverage was used. Contigs that were less than 500bp in length and with less than 80% high-quality base calls were eliminated from subsequent analysis.

Complementary dataset

We obtained 1669 assembled whole genomes of ST95 and associated metadata from Enterobase (117) accessed on March 12, 2020. Geographical and collection information of these whole genomes of ST95 is shown in Supplementary Table 1.

Bioinformatics analysis

Using the ResFinder database, we identified antimicrobial drug resistance (AMR) genes among the 80 clinical ST95 isolates' genomes (UTI and BSI) and among the 1669 assembled genomes from the Enterobase database (accessed 1 April 2020). We defined drug-resistant strains as those having at least one recognized drug resistance gene, and pan-susceptible strains as those not having any recognized drug resistance gene. Plasmid incompatibility group and replicon types were identified with the PlasmidFinder database (accessed April 2020). *FimH* types were identified with the FimTyper database (accessed April 2020). All identifications were done with Abricate version 1.0.1(118) with a 95% identity threshold across the reference sequences. Annotation was performed on all the assembled genomes with Prokka version 1.14.0 (119). Genes annotated as encoding hypothetical proteins were further analyzed on BLAST (accessed April 2021) (120). A pangenome of the entire data set was constructed with Roary version 3.13.0 (121) with a 95% identity cutoff. Here, the core genomes were defined as genes present in more than 95% of the collection of the isolates, and accessory genomes were defined as genes present in less than or equal to 95% of the collection. Of the accessory genomes, genes present in less than 15% of the collection were defined as cloud genes.

A concatenated core CDS alignment was made from the Roary output, and we extracted single nucleotide polymorphism (SNP) information with snp-sites with the default option (122). Phylogenetic trees were constructed with FastTree (123) with maximum likelihood method with the GTR model based on the SNP alignment and the presence and absence of accessory genes from the Roary output. Visualization was done with iToL version 6.1.2 (http://itol.embl.de).

Statistical analysis

All statistical analyses were conducted with R version 3.6.3 (63). We used the prcomp function for principal component analysis. The association of a specific gene with drug resistance was tested with multivariate analysis with the glm function adjusted for the carriage of UTI89 like plasmid and *fimH* types. For the regression analysis, only ST95 isolates from BSI and UTI patients were used to control for any other unmeasured potential confounding. We identified 473 BSI and UTI isolates based on the metadata information of Enterobase assessed on March 12, 2020. Combined with the 80 isolates from Northern California, 553 ST95 whole genome sequence data were used for this analysis. Relationships between each set of significantly associated genetic features were assessed with correlation coefficients. All graphs were plotted with ggplot2 (124).

Distribution of genes associated with pan-susceptibility

Once we identified ExPEC ST95 genes significantly associated with susceptibility to antimicrobial drugs, we analyzed the distribution and location of the gene in other ExPEC STs by BLAST (accessed June 2021) (120). We targeted ST131 because of their high frequency of drug resistance and the number of WGSs in the database (46, 110). We sought for these ST95 genes on (1) the ST131 chromosome, plasmid, or phage genes and (2) ST131 whole genome sequences deposited on NCBI database.

Accession numbers

The whole-genome shotgun sequence results described here have been deposited in DDBJ/ENA/GenBank under the accession numbers shown in Supplementary Table 4. Genome sequences have been deposited in NCBI BioProject under accession number PRJNA763994.

Results

Frequency and type of drug resistance genes among ST95 strains

We performed WGS analysis of 36 ST95 isolates that were previously collected from UTI patients in Northern California (108) and compared the sequences to those of 44 BSI ST95 isolates from Northern California (125) we previously reported, and 1669 ST95 WGSs deposited in the Enterobase database (117). We queried the carriage of recognized drug resistance genes against ResFinder and found that 837 (50%) of 1669 Enterobase ST95 isolates and 50 (62%) of 80 Northern California isolates did not contain any recognized resistance genes. Among the remaining 862 ST95 isolates, 112 recognized drug-resistant genes for 12 different classes of antibiotics were identified (Supplementary Table 2). Of these, 25 genes for 7 different classes of antibiotics were present in 80 clinical ST95 isolates from Northern California (Table 1). These classes of antimicrobial agents included aminoglycosides, beta-lactams, phenicols, trimethoprim, sulfonamide, tetracycline, and macrolides. Genes associated with fosfomycin, lincosamide, colistin, quinolone, and rifampicin resistance were not observed among any of the ST95 isolates from Northern California. In the period when the total annual submission of ST95 isolates to Enterobase was more than 50 (2010-2017), the prevalence of drug-resistant genes has remained relatively constant (33%-59%). When stratified by BSI or UTI, 86 (38%) of 225 BSI isolates and 132 (53%) of 248 UTI isolates were drug-resistant based on the ResFinder database.

Among the sequences from the 80 Northern California isolates, we found extended-spectrum beta-lactamase (ESBL) genes $bla_{CTX\text{-}M\text{-}14}$, $bla_{CTX\text{-}M\text{-}15}$, $bla_{TEM\text{-}1a}$, $bla_{TEM\text{-}1b}$, and $bla_{TEM\text{-}1c}$ in 1 (1.3 %), 1 (1.3 %), 3 (3.9 %), 21 (27 %), and 1 (1.3 %) of 80 isolates, respectively. The same set of ESBL genes was present in 29 (1.7 %), 44 (2.6 %), 18 (1.1 %), 335 (20.1 %), and 148 (8.9 %) of 1669 isolates among the sequences deposited in the Enterobase database, respectively. The temporal distribution of drug resistance genes in all the 1749 ST95 isolates is shown in Table 2.

Pan-genome analysis

A total of 35,134 genes constituted the pan-genome of 1749 ST95 WGSs. Of 35,134 genes, 3,688 (10.5%) were shared among more than 95% of the isolates (core genes) and 31,446 (89.5%) were distributed among subsets of the isolates (accessory genes). Of the latter, 29,522 genes were found in <15% of the isolates (cloud genes). Of the cloud genes, 13,970 (40%) genes were unique to only one isolate.

Phylogenetic and principal component analyses

Maximum-likelihood phylogenies of the SNP alignment of core genomes and the presence and absence of accessory genomes were obtained from FastTree (123) with the GTR model (Figure 1). Clustering based on *fimH* types was observed but we did not observe a strong correlation of drug resistance with phylogenetic trees based on the SNPs of core genomes and the presence and absence of accessory genes.

Principal component analysis (PCA) was also conducted to visually assess variables accounting for the variance. Cluster patterns based on the variance of SNPs and accessory genes were associated with the carriage of plasmid pUTI89* and with *fimH* types 18, 27, 30, 41, and 54 in core genomes and accessory genomes, respectively (Figure 2). However, the cluster patterns were not associated with drug resistance or pan-susceptibility.

Genes associated with drug-susceptible ST95 strains

We conducted a genome-wide analysis of ST95 genomes with 261 UTI ST95 isolates and 292 BSI ST95 isolates (80 from Northern California and 473 from the Enterobase ST95 collection) based on the metadata information. We limited the genome-wide association analysis to human BSI and UTI isolates (553 isolates in total) to control for disease status. The other isolates included those from nonhuman animals, the environment, and human non-urine or blood sources. The output of SNP-sites and Roary was used as variables. A logistic multivariate regression model adjusted for the carriage of pUTI89* and *fimH* types was conducted to identify genomic features significantly associated with drug susceptibility. We found 44 accessory genes significantly associated with strains that carried no recognized drug-resistance genes included in the ResFinder database (Table 3). We found 13 SNPs on core genes and 145 accessory genes associated with drug resistance (Supplementary Table 3). The correlation coefficient matrix of these significantly associated genes is shown in Figure 3.

Of the 44 genes associated with drug susceptible strains, 8 genes were annotated in the Prokka as *alkA* (DNA-3-methyladenine glycosylase 2), *cia* (colicin), *dinG*_1 (3'-5' exonuclease), *livK* (ABC transporter substrate-binding protein LivK), *maX_3* (PTS sugar transporter subunit IIA), *mltC* (membrane-bound lytic murein transglycosylase MltC), *traD* (type IV conjugative transfer system coupling protein TraD), and *umuD_2* (translesion error-prone DNA polymerase V autoproteolytic subunit). The other 36 genes were annotated by Prokka as encoding hypothetical proteins of unknown function. The sequences of 36 hypothetical proteins were sought by BLAST and information on protein function was obtained. Of these, 30 proteins were annotated in the NCBI database. Eight genes were annotated to encode transfer system or transporter-related proteins and 5 genes were annotated as phage genes. The functions of 44 genes are shown in Table 3.

Distribution of genes associated with pan-susceptible ST95 strains

Of 44 accessory genes significantly associated with drug susceptibility, 13 were found on pUTI89*, 19 were on non-pUTI89* plasmids, and 15 were on the chromosome (Table 3). Three genes were found both on pUTI89* plasmid and chromosomes.

We also sought these 44 gene sequences in ExPEC strains belonging to ST131, which are largely resistant to one or more antimicrobial agents. Fifteen of 44 genes did not match with any of the ST131 WGSs deposited in the NCBI database as of July 28, 2021 (Table 3). The other 29 genes had a match ranging from 1 to 28 from 7458 *Escherichia coli* O25b:H4-ST131 WGSs. NCBI has 127 complete whole genome sequences of *Escherichia coli* O25b:H4-ST131 and 7331 plasmid or short genome sequences of *Escherichia coli* O25b:H4-ST131 deposited. Of these sequences, 4 complete chromosome and 30 plasmid sequences carried a range of 1 to 14 genes associated with the pan-susceptible ST95 strains (Table 4).

Discussions

We identified genomic features associated with drug susceptibility of ExPEC ST95 lineage isolated from BSI and UTI patients, as well as other sources. We analyzed 1669 ST95 isolates deposited in Enterobase (as of March 12, 2020) and 80 ST95 isolates from UTI or BSI patients in Northern California. We also conducted a genome-wide association study with 553 ST95 BSI and UTI human isolates. The sources of other isolates submitted to Enterobase include soil, water, vegetables, animals other than humans, and human samples other than urine or blood (e.g., stool or wound). Our previous study of WGSs of 86 isolates of ST95 from BSI patients in the United States found that the carriage of UTI89 like plasmid (pUTI89*) and *fimH6* was significantly associated with drug susceptibility (5). We also reported that in Northern California, ST95 strains remained persistently susceptible to most of the drugs we tested despite being one of the most prevalent ExPEC STs isolated from patients with community-acquired UTI (108). We hypothesized that ST95 strains restrict the acquisition of mobile DNA elements carrying drug-resistance genes and yet maintain a competitive advantage without becoming drug-resistant among the human intestinal microbiota.

Of the 1749 ST95 WGSs, 862 (49%) carried one or more drug-resistance genes. Among these, 94 (36%) of 261 UTI ST95 isolates and 154 (53%) of 292 BSI ST95 isolates carried drug-resistance genes, respectively. We observed a lower prevalence of drug-resistant ST95 isolates in Northern California compared to those in the Enterobase database (38% and 50%, respectively). This may be due to a submission bias to Enterobase—for example, drug-resistant ST95 may have been more likely to be sequenced and deposited in Enterobase. The Northern California isolates represented population-based samples (97, 113).

Principal component analysis and phylogenetic analysis based on the SNPs on core genes and the presence and absence of accessory genes were conducted to identify potential confounding factors associated with clustering. In both analyses, carriage of pUTI89* and *fimH* types were strongly associated with clustering. This suggests that the carriage of pUTI89* and the *fimH* types explain the majority of genomic variation we extracted from the pangenome analysis of the ST95 WGSs. This is consistent with our previous findings (5), and also offers more evidence for *fimH* typing as a useful method for genotyping ExPEC (126).

We used multivariate logistic models to identify genomic features significantly associated with drug pan-susceptibility while controlling for *fimH* types, and the carriage of pUTI89*. We limited the whole genome sequence data to human BSI or UTI isolates (553 isolates in total) to control for disease status. We found 44 accessory genes associated with drug susceptibility. We did not identify any SNPs in the core genes significantly associated with drug-susceptible strains. Of these 44 genes, 15 were not found in any of 7458 *E. coli* ST131 WGSs deposited in the NCBI nucleotide database. ST131 strains are frequently associated with multidrug resistance (4, 45),

and thus these 15 ST95 genes not found in ST131 may include genetic features that confer upon ST95 strains the ability to resist becoming drug-resistant. The other 29 genes showed 1 to 28 matches with the ST131 genome sequences, but these matches represented less than 1% of the ST131 WGSs, whereas they represented 10-87% of the ST95 WGSs (Table 3).

Annotation of these 15 genes not found on any of the ST131 in the NCBI nucleotide database revealed that two genes were annotated as PTS sugar transporter subunit IIA and IIB, respectively. Previous reports revealed that PTS sugar transporters are involved in the penetration of polar antibiotics (127, 128) Three genes were associated with phage. The remaining functions were colicin, malate dehydrogenase, DNA breaking-rejoining protein, surface exclusion protein, and peptidase M14. Five were hypothetical proteins, functions unknown.

Acquisition and maintenance of accessory genes by *E. coli* must provide a population of *E. coli* strains an adaptive advantage in an ecological niche. ST95 strains may have evolved compensatory niche-adaptive mechanisms to remain viable in a variety of ecological niches (human and other animal intestines, environment) despite remaining susceptible to antibiotics. One such mechanism is the role of amino acid transporters reported to contribute to metabolic adaptation of uropathogenic *E. coli* (UPEC) in the urinary tract (129).

A limitation of this study is that the drug resistance phenotype of 1669 ST95 WGS obtained from the Enterobase database was predicted from the presence of drug resistance genes annotated in ResFinder and not by any experimental phenotyping tests. Although genotypic resistance has been shown to have 100% sensitivity and specificity in predicting phenotypic resistance in *E. coli* by some studies (130, 131), we cannot rule out the possibility of misclassifying drug resistance in this study. Nevertheless, we found strong associations of multiple accessory genes with the absence of recognized drug-resistance genes in ST95 strains.

Another limitation is the lack of consistent meta-data in the Enterobase database. Although the Enterobase database had data on 1669 ST95 isolates, the majority of them lacked information on the source of ST95 strains, which led to the reduction of sample size for the genome-wide association study (n=553). Lastly, submission bias of ST95 to Enterobase could have affected the prevalence of drug-resistant strains in our collection. The prevalence of drug-resistant ST95 strains in this collection may over-estimate the actual prevalence of drug-resistant ST95 in the human population.

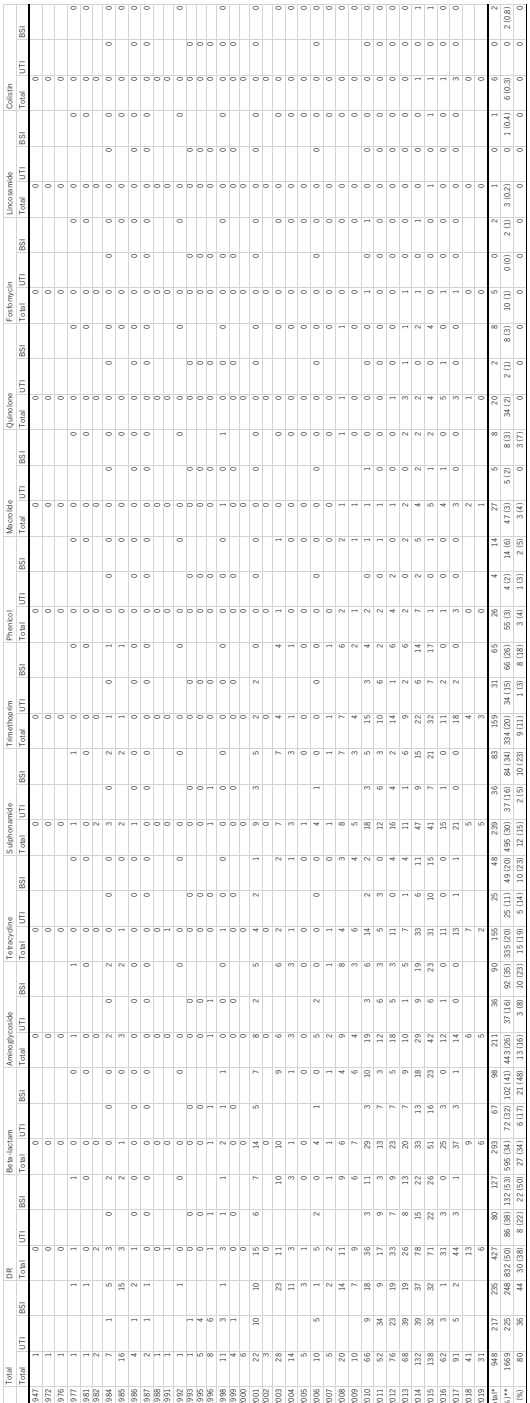
In conclusion, we identified several genes associated with drug pan-susceptibility in ExPEC ST95 strains, which suggests that the ST95 lineage may have evolved to compensate for its susceptibility to antibiotics by acquiring unique fitness genes that enable a subpopulation of them to survive in multiple environmental niches without having to gain drug resistance. Experimental confirmation of these findings is needed to support this proposal, which is feasible with the relatively small number of unique genes in pan-susceptible ST95 strains we identified.

Tables

Table 1: Frequency of antimicrobial resistance genes among *E. coli* isolates belonging to ST95 registered in Enterobase database and obtained in Northern California respectively, identified from ResFinder

Antibiotics class	Resistance gene name	Enterobase Total (%)	North California Total (%)
Pan susceptible		837 (50)	50 (62)
Aminoglycoside	aac(3)-IId_1	36 (2.16)	2 (2.56)
	aadA2_1	39 (2.34)	2 (2.56)
	aadA5_1	33 (1.98)	4 (5.13)
	ant(2")-Ia_1	2 (0.12)	2 (2.56)
	ant(3")-Ia_1	38 (2.28)	1 (1.28)
	aph(3")-Ib_5	306 (18.33)	6 (7.69)
	aph(6)-Id_1	254 (15.22)	7 (8.97)
Beta-lactam	blaCTX-M-14_1	29 (1.74)	1 (1.28)
	blaCTX-M-15_1	44 (2.64)	1 (1.28)
	blaTEM-1A_1	18 (1.08)	3 (3.85)
	blaTEM-1B_1	335 (20.07)	21 (26.92)
	blaTEM-1C_1	148 (8.87)	1 (1.28)
Phenicol	catA1_1	16 (0.96)	2 (2.56)
	floR_2	17 (1.02)	1 (1.28)
Trimethoprim	dfrA12_8	33 (1.98)	2 (2.56)
	dfrA17_1	63 (3.77)	5 (6.41)
	dfrA5_1	153 (9.17)	2 (2.56)
Sulphonamide	sul1_2	1 (0.06)	1 (1.28)
	sul1_5	107 (6.41)	5 (6.41)
	sul2_2	278 (16.66)	4 (5.13)
	sul2_3	114 (6.83)	3 (3.85)
Tetracycline	tet(A)_6	211 (12.64)	8 (10.26)
	tet(B)_1	1 (0.06)	1 (1.28)
	tet(B)_2	107 (6.41)	5 (6.41)
Macrolide	mph(A)_2	42 (2.52)	3 (3.85)

Table 2:



Year	Total			DR			Beta-lactam			Aminoglycoside			Tetracycline			Sulphonamide			Trimethoprim			Phenicol			Macrolide			Quinolone			Fosfomycin			Lincosamide			Colistin		
	Total	UTI	BSI	Total	UTI	BSI	Total	UTI	BSI	Total	UTI	BSI	Total	UTI	BSI	Total	UTI	BSI	Total	UTI	BSI	Total	UTI	BSI	Total	UTI	BSI	Total	UTI	BSI	Total	UTI	BSI	Total	UTI	BSI
1947	1			0			0			0						0			0			0			0			0			0			0			0	
1972	1			0			0			0						0			0			0			0			0			0			0			0	
1976	1			0			0			0						0			0			0			0			0			0			0			0	
1977	1		1	0			0			0			1	0		0			0			0			0			0			0			0			0	
1981	1		1	0			0			0			0			0			0			0			0			0			0			0			0	
1982	2			2			0			0						2			0			0			0			0			0			0			0	
1984	7	1	5	3	0	2	0	0	0	2	0	2	0	0	0	3	0	2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1985	16		15	3		2	1		1	3		2				2		2	1		1	0		0	0		0	0		0	0		0	0		0	0	
1986	4	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1987	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1988	1			0			0			0			0			0			0			0			0			0			0			0			0	
1991	1			1			0			0			1			0			0			0			0			0			0			0			0	
1992	1		1	0		0	0		0	0		0	0		0	0		0	0		0	0		0	0		0	0		0	0		0	0		0	0	
1993	1			0		0	0		0	0		0	0		0	0		0	0		0	0		0	0		0	0		0	0		0	0		0	0	
1995	5	1	4	0		0	0		0	0		0	0		0	0		0	0		0	0		0	0		0	0		0	0		0	0		0	0	
1996	8	6	2	1		1	1		1	1		1	0		0	1		1	0		0	0		0	0		0	0		0	0		0	0		0	0	
1998	11	3	1	3	1	1	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
1999	4	1		0			0			0			0			0			0			0			0			0			0			0			0	
2000	6			0			0			0			0			0			0			0			0			0			0			0			0	
2001	22	10	10	15	6	7	14	5	7	8	2	5	4	2	1	9	3	5	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2002	3			0			0			0			0			0			0			0			0			0			0			0			0	
2003	28		23	11		10	10		9	6		6	2		2	7		7	4		4	1		1	0		0	0		0	0		0	0		0	0	
2004	14		11	3		3	1		3	3		3	1		1	3		3	1		1	0		0	0		0	0		0	0		0	0		0	0	
2005	5		3	1		0	0		0	0		0	0		0	1		0	0		0	0		0	0		0	0		0	0		0	0		0	0	
2006	10	5	1	5	2	0	4	1	0	5	2	0	0	0	0	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2007	5		2	2		1	1		1	2		1	1		0	1		1	1		1	0		0	0		0	0		0	0		0	0		0	0	
2008	20		14	11		9	6		4	9		8	4		3	8		7	7		6	2		2	1		1	0		0	0		0	0		0	0	
2009	10		7	9		6	7		6	4		4	3		3	4		4	5		5	2	1	1	0		0	0		0	0		0	0		0	0	
2010	66	9	18	39	3	11	29	3	10	19	3	6	14	2	2	18	3	5	15	3	4	2	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0
2011	52	34	9	17	9	3	13	7	3	12	6	3	5	3	0	12	6	3	10	6	2	2	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0
2012	76	23	19	33	7	9	23	7	5	18	5	3	11	0	4	16	4	2	14	1	6	4	2	0	1	1	0	0	0	0	0	0	0	0	0	0	0
2013	68	39	19	26	8	13	20	7	9	10	1	5	7	1	4	11	1	6	9	2	6	2	0	2	2	0	2	3	1	1	1	0	0	0	0	0	0
2014	132	39	37	78	15	22	33	13	18	29	9	19	33	6	11	47	9	15	22	6	14	7	2	5	4	2	2	2	0	2	1	0	1	0	0	1	0	1
2015	138	32	32	71	22	26	51	16	23	42	6	23	31	10	15	41	7	21	32	7	17	1	0	1	5	1	2	4	0	4	0	0	1	0	1	1	0	1
2016	62	3	1	31	3	0	25	3	0	12	1	0	11	0	0	15	1	0	11	2	0	1	0	0	4	1	0	5	1	0	1	0	0	0	0	1	0	0
2017	91	5	2	44	3	1	37	3	1	14	0	0	13	1	1	21	0	0	18	2	0	3	0	0	3	0	0	3	0	0	1	0	0	0	0	3	0	0
2018	41			13			9			6			7			5			4			0			2			1			0			0			0	
2019	31			6			6			5			2			5			3			0			1			0			0			0			0	
Total	948	217	295	422	90	127	470	47	98	213	36	90	166	25	46	269	36	83	169	31	45	55	4	14	77	5	8	40	2	8	5	0	0	2	1	4	1	2
Total (%)**	1466	225	248	832 (57)	86 (38)	132 (53)	595 (41)	72 (32)	102 (41)	443 (30)	37 (16)	92 (36)	335 (23)	25 (11)	49 (20)	496 (33)	37 (16)	84 (34)	334 (23)	66 (29)	55 (3)	4 (2)	14 (6)	47 (3)	5 (2)	8 (3)	34 (2)	2 (1)	8 (3)	10 (1)	0 (0)	2 (1)	3 (0.2)	1 (0.4)	6 (0.3)	2 (0.8)	
In CA (%)	801	36	44	30 (38)	8 (22)	22 (50)	27 (34)	6 (17)	21 (48)	13 (16)	3 (8)	10 (23)	15 (19)	5 (14)	10 (23)	12 (15)	2 (5)	10 (23)	9 (11)	1 (3)	8 (18)	3 (4)	1 (3)	2 (5)	3 (4)	0	3 (7)	0	0	0	0	0	0	0	0	0	0	0

*isolates with collection year information

**includes Enterobase isolates without collection year information

DR: drug resistant, CA: California

Table 3: Genes significantly associated with ST95 strains lacking drug-resistance genes.

			UTI		BSI			
Gene/SNP	Annotation	location	OR [95% CI]	P-value	OR [95% CI]	P-value	ST131	ST95 (%)
Pan-Susceptible								
alkA	DNA-3-methyladenine glycosylase 2	chromosome	5.5 [1.7-20.4]	0.007	2.5 [1-6.6]	0.046	5	1420 (80)
cia	colicin	plasmid	4.8 [1.8-13.9]	0.003	2.6 [1.1-6.2]	0.023	0	318 (18)
dinG_1	3'-5' exonuclease	plasmid	5.3 [2-15.4]	0.001	4.7 [1.9-12.3]	0.001	1	311 (18)
livK	leucine-specific binding protein precursor, high-affinity branched-chain amino acid ABC transporter substrate-binding protein LivK	chromosome	3.8 [1.2-13.5]	0.03	4 [1.5-12]	0.008	5	1546 (87)
manX_3	PTS sugar transporter subunit IIA	chromosome	3.8 [1.2-13.5]	0.03	3.9 [1.5-11.8]	0.009	0	1547 (87)
mltC	membrane-bound lytic murein transglycosylase MltC	chromosome	2.1 [1-4.4]	0.042	2.1 [1-4.4]	0.043	1	1336 (75)
traD	type IV conjugative transfer system coupling protein TraD	plasmid	4.1 [1.9-9.2]	0	2.2 [1.2-4.3]	0.014	1	760 (43)
umuD_2	translesion error-prone DNA polymerase V autoproteolytic subunit	pUTI89	11.2 [3.6-38.5]	0	4.2 [1.6-10.9]	0.003	1	371 (21)
group_406	conjugal transfer mating pair stabilization protein TraG	plasmid	6.1 [2.5-16.1]	0	4.5 [2.2-9.7]	0	22	379 (21)
group_1304	hypothetical protein	pUTI89	43.5 [7.9-1000]	0	9.7 [2.5-62.5]	0.004	28	196 (11)
group_2195	DUF4113 domain-containing protein	plasmid	12 [4-40]	0	6.7 [1.9-24.4]	0.003	0	462 (26)
group_2212	DNA repair protein UmuC	pUTI89	18.9 [5.5-90.9]	0	5.8 [2-19.6]	0.002	2	264 (15)
group_2341	incFII family plasmid replication initiator RepA	plasmid	3.9 [1.6-10.2]	0.004	3.4 [1.5-7.5]	0.003	25	377 (21)
group_2492	conjugal transfer pilus acetylase TraX	plasmid	4.9 [2-12.7]	0.001	6.8 [3.3-15.2]	0	25	423 (24)
group_2899	phage integrase family protein	pUTI89	21.7 [6.2-83.3]	0	6.7 [1.9-24.4]	0.003	28	352 (20)
group_3175	DUF945 domain-containing protein	pUTI89	6.6 [1.7-43.5]	0.017	4.1 [1.1-19.6]	0.048	23	191 (11)
group_3361	transglycosylase SLT domain-containing protein	pUTI89	6.6 [1.9-30.3]	0.006	3.3 [1.1-11.2]	0.044	23	218 (12)
group_3450	fertility inhibition protein FinO	plasmid	3.4 [1.4-8.8]	0.009	3 [1.4-6.4]	0.005	25	384 (22)
group_3636	type IV conjugative transfer system protein TraL	pUTI89	6.2 [1.6-41.7]	0.022	5.6 [1.3-38.5]	0.036	22	169 (10)
group_4084	phage tail protein	chromosome	6.3 [1.2-50]	0.041	4.2 [1.4-13.3]	0.012	1	311 (18)
group_4555	TrbJ conjugal transfer protein	plasmid	6.4 [2.5-17.5]	0	4.5 [2.1-9.7]	0	12	395 (22)
group_5037	IS21 family transposase	pUTI89	23.3 [4.3-500]	0.003	9.3 [2.3-62.5]	0.005	5	191 (11)
group_5504	nuclease-like protein	plasmid	4.8 [2-12.2]	0.001	4.7 [2.4-9.6]	0	9	495 (28)
group_6014	IncF conjugal transfer surface exclusion protein TraT	plasmid	4 [1.7-9.4]	0.001	4.5 [2.2-9.3]	0	23	446 (25)
group_6737	phage minor tail protein G	chromosome	2.6 [1.2-5.9]	0.015	2.5 [1.2-5.3]	0.018	0	795 (45)
group_6764	hypothetical protein	plasmid	4.8 [1.6-17.5]	0.008	4 [1.3-15.6]	0.027	28	187 (11)
group_7168	tail fiber domain-containing protein	chromosome	6.3 [1.2-50]	0.041	4.3 [1.4-15.2]	0.016	1	305 (17)
group_8375	ISL3-like element ISEc53 family transposase	pUTI89, chromosome	16.4 [4.1-111.1]	0	8.9 [2.5-41.7]	0.002	4	238 (13)
group_8556	phage portal protein	chromosome	2.6 [1.2-5.9]	0.015	2.5 [1.2-5.4]	0.014	0	603 (34)
group_9933	DinI-like family protein	pUTI89, chromosome	10.8 [3.3-38.5]	0	4.8 [1.7-14.1]	0.003	1	373 (21)
group_9939	malate dehydrogenase	plasmid	3 [1.3-7.2]	0.011	4.2 [2-9.3]	0	0	400 (23)
group_9942	hypothetical protein	plasmid	4.1 [1.8-10]	0.001	3.8 [1.9-7.8]	0	0	459 (26)
group_9945	hypothetical protein	plasmid	4.1 [1.8-10.2]	0.001	3.3 [1.5-7.5]	0.003	0	382 (22)
group_10528	ATP-binding protein	pUTI89	19.6 [3.7-333.3]	0.005	16.1 [3-333.3]	0.009	1	183 (10)
group_11645	DNA breaking-rejoining protein	chromosome	2.6 [1.2-5.9]	0.015	2.5 [1.2-5.3]	0.018	0	610 (34)
group_12425	PAS domain-containing protein	pUTI89	4.5 [1.4-17.2]	0.016	9.2 [2.2-62.5]	0.007	7	226 (13)
group_13166	PTS sugar transporter subunit IIB	chromosome	3.8 [1.2-13.5]	0.03	3.7 [1.5-10.5]	0.009	0	1545 (87)
group_13675	ClbS/DfsB family four-helix bundle protein	pUTI89, chromosome	10 [2.9-47.6]	0.001	9.3 [2.7-43.5]	0.001	4	246 (14)
group_14237	surface exclusion protein	plasmid	3.1 [1.4-7]	0.005	4.9 [2.5-10.1]	0	0	528 (30)
group_14238	phage tail assembly protein T	chromosome	2.6 [1.2-6]	0.017	2.3 [1.1-5.2]	0.033	0	542 (31)
group_14249	hypothetical protein	plasmid	3.7 [1.6-8.8]	0.002	4.5 [2.2-9.2]	0	0	485 (27)
group_15802	replication regulatory protein RepA	plasmid	3.2 [1.4-7.7]	0.009	5.2 [2.6-10.8]	0	4	522 (29)
group_17896	DUF2190 family protein	chromosome	2.6 [1.2-5.9]	0.015	2.5 [1.2-5.4]	0.014	0	597 (34)
group_20939	peptidase M14	plasmid	3.2 [1.4-7.7]	0.006	4.5 [2.2-9.2]	0	0	485 (27)

Annotation, location, and odds ratio with 95% confidence interval are shown for UTI and BSI isolates. These ST95 genes that had a matching sequence in *Escherichia coli* O25b:H4-ST131 WGSs from NCBI database are shown. Prevalence of these genes among 1774 ST95 collection is shown.

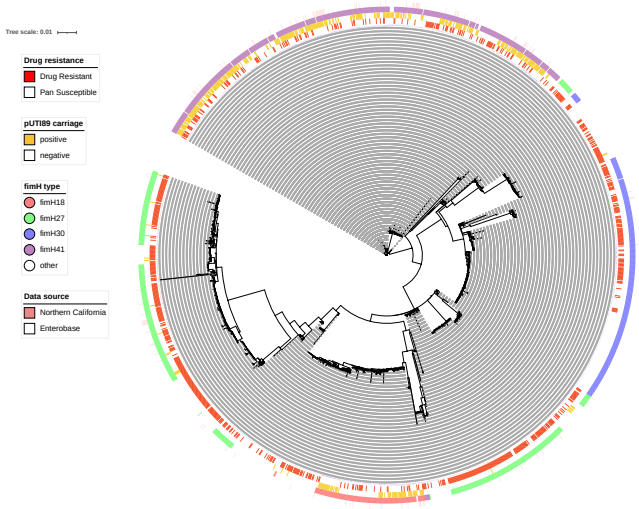
Table 4: NCBI accession number and the number of matching regions of ST131 whole genome sequence that carries one or more genes of 44 ST95 genes found to be associated with ST95 strains lacking drug-resistance genes.

Location	NCBI accession number	Number of genes found
ST131 chromosome	CP063774	11
	CP051615	5
	CP051609	5
	HG941718	9
ST131 plasmid	AP018456	12
	AP018457	11
	AP018458	12
	AP019526	9
	CP063775	3
	CP051616	12
	CP051610	9
	CP015086	1
	HG941719	13
	JX077110	7
	MK295817	9
	MK295818	11
	MK295819	11
	MK295820	12
	MK295821	14
	MK295822	5
	MK295823	14
	MK295824	11
	MK295825	12
	MK295826	6
	MK295827	7
	MK295828	12
	MK295829	12
	MK295830	12
	MK295831	14
	MK295832	11
	MK295833	14
	MK295834	12
	MW646300	13
	MW646306	7

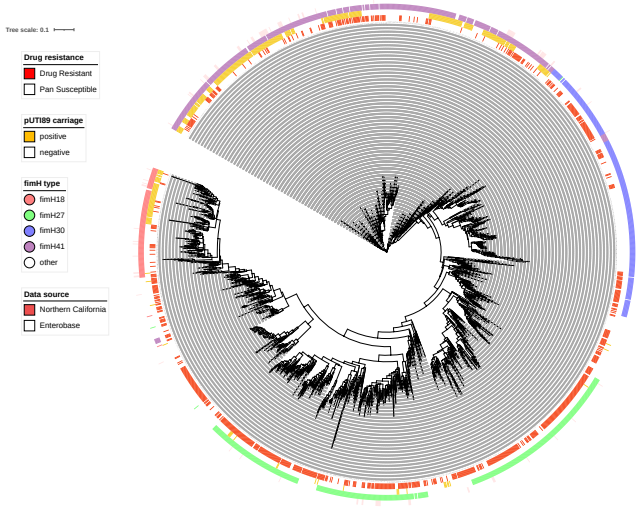
Figures

Figure 1: Maximum-likelihood phylogeny of 1774 ST95 isolates constructed with FastTree and visualized with iTOL

A



B



A: Phylogeny based on the core genome

B: Phylogeny based on presence and absence of accessory genome

Figure 2: Principle component analysis of 1774 ST95 isolates based on (A) core genome and (B) presence and absence of accessory genome

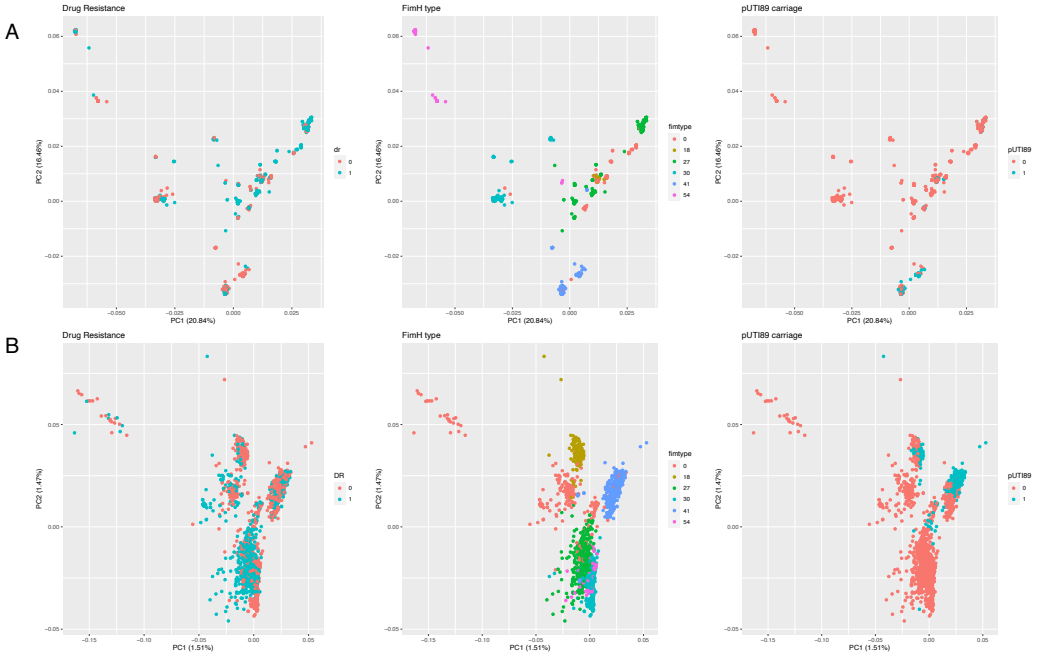
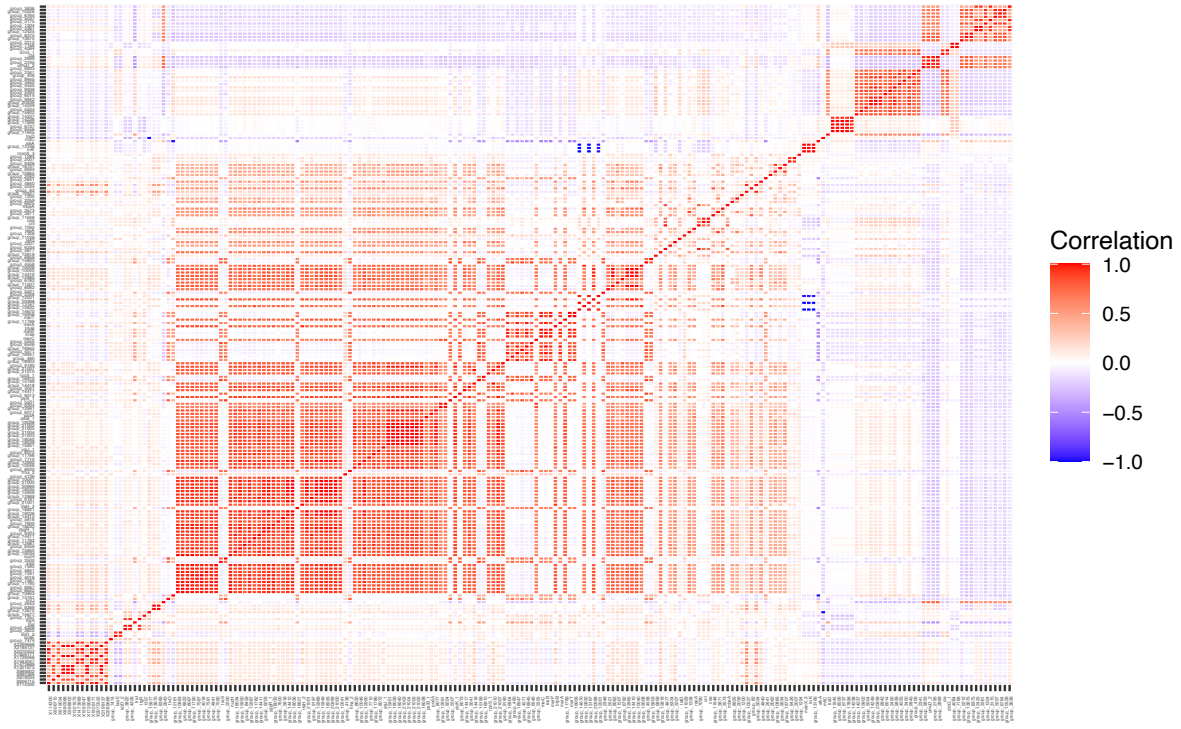


Figure 3: The correlation coefficient matrix of ST95 genes significantly associated with ST95 strains lacking drug-resistance genes.



Chapter 5. Conclusion

5.1. Summary

This dissertation aimed to characterize host and microbial features associated with AMR ExPEC infections. Since AMR UTI and BSI largely result from infection by an *E. coli* strain that originates in the infected person's intestine, we first described host-related demographic and behavioral risk factors for carriage of AMR *E. coli*. In Chapter 2, we identified risk factors associated with the carriage of drug resistant *E. coli* in the gut by literature review and meta-analysis from 15 studies published from 2014 to 2019. In Chapter 3, we conducted a cross-sectional survey at University of California, Berkeley community targeting healthy adults to identify risk factors related to dietary habit that affect the probability of carrying multidrug resistant *E. coli* in their gut. Lastly, we applied laboratory and computational biology tools to investigate a pandemic ExPEC lineage ST95, which is globally disseminated but remains susceptible to most of the drugs used to treat UTI and BSI. Much of the literature examining risk factors for AMR bacterial infections focus on host-related factors. Chapter 4 was an attempt to identify microbial factors associated not with AMR but with pan-susceptibility in order to characterize counterfactual microbe-related risk factors for AMR.

The limitation throughout this dissertation work is that all identified risk factors are based on correlation analyses and therefore do not provide any evidence for causal relationships. Nevertheless, in Chapter 2 and Chapter 3, identification of factors significantly associated with the carriage of drug-resistant *E. coli* will still be useful for identifying individuals with high risk to implement early focused interventions. In the following sections, brief summary of each studies and the final conclusions are discussed.

5.2. Literature review and meta-analysis of risk factors associated with fecal carriage of drug resistant *Escherichia coli*

In this literature review and meta-analysis, we found that previous antimicrobial treatment, diarrhea, travel to India, and vegetarian diet was identified to significantly increase the risk of fecal carriage of drug resistant *E. coli*. Interestingly, regardless of the excessive antibiotics use in clinical situations and food animal husbandry in the US, no literature from the US met the inclusion criteria for this literature review. This highlighted the importance of conducting such studies or establishing a surveillance system in the US to systematically collect data to identify risk factors associated with fecal carriage of drug resistant genes or bacteria in North American communities.

5.3 Cross-sectional study of risk factors associated with fecal carriage of multidrug resistant *Escherichia coli*

In Chapter 3, we conducted a cross sectional survey of healthy volunteers from a university

community at University of California, Berkeley. From the literature review, we found that, as of 2019, no study has been reported from any North American country that investigated risk factors for fecal carriage of drug resistant genes or drug-resistant Gram-negative bacteria . Here, we identified red meat consumption to be associated with the increased risk, and surprisingly, fish consumption to be associated with decreased risk for the carriage of multidrug resistant *E. coli*, adjusted for biological sex, employment status, frequently used supermarket and previous travel. The major limitation of this study was the small sample size of the study subjects, and that the surveillance relied on self-reports of the participants. Thus, further studies with a larger population and at other locations will be essential for establishing the generalizability of our findings.

5.4 Genome wide association study of ExPEC ST95

In Chapter 4, we aimed to identify genomic features associated with the pan-susceptibility of ExPEC ST95. The reason we focused on ST95 is because of its uniqueness in the pattern of drug resistance. Despite ST95 being one of the most common causes of UTI and BSI, the majority of ST95 isolates remain pan-susceptible to most of the drugs used to treat these infections. That is, this strain must frequently come under the selective pressure of antimicrobial agents to become resistant, and yet, unlike the other common ExPEC lineages, it has remained relatively resistant to becoming AMR.

We analyzed 1749 whole genome sequences of ST95 obtained from Northern California community and Enterobase database and found 44 accessory genes associated with absence of carriage of recognized drug-resistance genes. The limitation of this genome-wide association study is we lack experimental confirmation of these findings. However, these findings will pave the way to devise such experimental studies, which will be carried out in the Riley Lab at UC Berkeley.

5.5 Final considerations

The ultimate goal of the studies described in this dissertation is to generate new ideas that can be used to devise novel strategies to combat antimicrobial resistant infections caused by Gram-negative bacterial pathogens, which continues to increase in incidence globally. To achieve this objective, we were able to determine that certain dietary behavior and travel destinations can contribute to colonization with drug-resistant *E. coli*. Such knowledge could be potentially used to devise prevention of acquisition of drug-resistant *E. coli*. We also approached this objective from a different perspective, which was to find potential bacterial mechanisms to resist acquiring drug resistant genes. Such knowledge could be exploited to devise biologic interventions to interrupt *E. coli* from gaining drug-resistance genes.

These studies have provided a foundation for further investigations that include experimental studies as well as additional studies in other communities to assess generalizability of our findings. In conclusion, this dissertation work highlights the importance of global bacterial drug resistant infectious disease problem and opened a path towards further studies that may ultimately lead to the creation of new prevention strategies against AMR infections. It may also have opened a new door to characterize and understand how bacteria (nature) have figured out a way to prevent becoming AMR.

Bibliography

1. Kaper JB, Nataro JP, Mobley HLT. 2004. Pathogenic Escherichia coli. Nat Rev Microbiol 2:123–140.
2. Köhler CD, Dobrindt U. 2011. What defines extraintestinal pathogenic Escherichia coli? Int J Med Microbiol 301:642–647.
3. Laupland KB. 2013. Incidence of bloodstream infection: A review of population-based studies. Clin Microbiol Infect.
4. Riley LW. 2014. Pandemic lineages of extraintestinal pathogenic Escherichia coli. Clin Microbiol Infec 20:380–390.
5. Stephens CM, Adams-sapper S, Sekhon M, Johnson JR, Riley LW. 2017. Genomic Analysis of Factors Associated with Low Prevalence of Antibiotic Resistance in Extraintestinal Pathogenic Escherichia coli Sequence Type 95 Strains. mSphere 2:e00390-16.
6. Johnson JR, Russo TA. 2002. Extraintestinal pathogenic Escherichia coli: "The other bad E coli." J Lab Clin Med 139:155–162.
7. Flores-Mireles AL, Walker JN, Caparon M, Hultgren SJ. 2015. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat Rev Microbiol 13:269–284.
8. Maluta RP, Logue CM, Casas MRT, Meng T, Guastalli EAL, Rojas TCG, Montelli AC, Sadatsune T, Ramos MDC, Nolan LK, Silveira WD Da. 2014. Overlapped sequence types (STs) and serogroups of avian pathogenic (APEC) and human extra-intestinal pathogenic (ExPEC) Escherichia coli isolated in Brazil. PLoS One 9.
9. Manges AR, Geum HM, Guo A, Edens TJ, Fibke CD, Pitout JDD. 2019. Global Extraintestinal Pathogenic Escherichia coli (ExPEC) Lineages 32:1–25.
10. Magruder M, Sholi AN, Gong C, Zhang L, Edusei E, Huang J, Albakry S, Satlin MJ, Westblade LF, Crawford C, Dadhania DM, Lubetzky M, Taur Y, Littman E, Ling L, Burnham P, De Vlaminck I, Pamer E, Suthanthiran M, Lee JR. 2019. Gut uropathogen abundance is a risk factor for development of bacteriuria and urinary tract infection. Nat Commun 10:1–9.
11. Riley LW. 2020. Extraintestinal Foodborne Pathogens. Annu Rev ofFood Sci Technol 11:1–20.
12. Griffin PM, Manges AR, Johnson JR. 2012. Food-borne origins of escherichia coli causing extraintestinal infections. Clin Infect Dis 55:712–719.
13. Manges AR, Johnson JR. 2015. Reservoirs of Extraintestinal Pathogenic Escherichia coli. Microbiol Spectr 3:1–12.
14. Matsui Y, Hu Y, Rubin J, de Assis RS, Suh J, Riley LW. 2020. Multilocus sequence

typing of Escherichia coli isolates from urinary tract infection patients and from fecal samples of healthy subjects in a college community. Microbiologyopen 9:1225–1233.
15. Chen Y, Chang C, Chiu THT, Lin M, Lin C. 2020. The risk of urinary tract infection in vegetarians and non-vegetarians : a prospective study. Sci Rep 1–9.
16. Centers for Disease Control and Prevention. 2013. Antibiotic resistance threats in the United States, 2013.
17. WHO (World Health Organisation). 2014. ANTIMICROBIAL RESISTANCE Global Report on Surveillance.
18. Kaye KS, Pogue JM. 2015. Infections Caused by Resistant Gram-Negative Bacteria: Epidemiology and Management. Pharmacotherapy 35:949–962.
19. Colello R, Etcheverría AI, Di Conza JA, Gutkind GO, Padola NL. 2015. Antibiotic resistance and integrons in shiga toxin-producing escherichia coli (STEC). Brazilian J Microbiol 46:1–5.
20. Begum J, Mir NA, Dev K, Khan IA. 2018. Dynamics of antibiotic resistance with special reference to Shiga toxin-producing Escherichia coli infections. J Appl Microbiol 125:1228–1237.
21. Centers for Disease Control and Prevention. 2017. E. coli (Escherichia coli) Prevention.
22. Thaden JT, Jr VGF, Sexton DJ, Anderson DJ. 2016. Increasing Incidence of Extended-Spectrum β-Lactamase- Producing Escherichia coli in Community Hospitals throughout the Southeastern United States. Infect Control Hosp Epidemiol 37:49–54.
23. Allocati N, Masulli M, Alexeyev MF, Di Ilio C. 2013. Escherichia coli in Europe: An overview. Int J Environ Res Public Health 10:6235–6254.
24. Kaushik M, Kumar S, Kapoor RK, Gulati P. 2019. Integrons and antibiotic resistance genes in water-borne pathogens: Threat detection and risk assessment. J Med Microbiol 68:679–692.
25. Saliu EM, Vahjen W, Zentek J. 2017. Types and prevalence of extended-spectrum beta-lactamase producing Enterobacteriaceae in poultry. Anim Heal Res Rev 18:46–57.
26. Yamaji R, Friedman CR, Rubin J, Suh J, Thys E, McDermott P, Hung-Fan M, Riley LW. 2018. A Population-Based Surveillance Study of Shared Genotypes of Escherichia coli Isolates from Retail Meat and Suspected Cases of Urinary Tract Infections . mSphere 3:1–12.
27. Raphael E, Wong LK, Riley LW. 2011. Extended-spectrum beta-lactamase gene sequences in Gram-negative saprophytes on retail organic and nonorganic spinach. Appl Environ Microbiol 77:1601–1607.
28. Sapkota S, Adhikari S, Pandey A, Khadka S, Adhikari M, Kandel H, Pathak S, Pandey A. 2019. Multi-drug resistant extended-spectrum beta-lactamase producing E. coli and

Salmonella on raw vegetable salads served at hotels and restaurants in Bharatpur, Nepal. BMC Res Notes 12:516.

29. Weiner LM, Webb AK, Limbago B, Dudeck MA, Patel J, Kallen AJ, Edwards JR, Sievert DM. 2016. Antimicrobial-Resistant Pathogens Associated with Healthcare-Associated Infections: Summary of Data Reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2011-2014. Infect Control Hosp Epidemiol 37:1288–1301.
30. Donskey CJ. 2004. The role of the intestinal tract as a source for transmission of nosocomial pathogens. Clin Infect Dis 39:219–226.
31. Carlet J. 2012. The gut is the epicentre of antibiotic resistance. Antimicrob Resist Infect Control 1:1.
32. Woerther PL, Burdet C, Chachaty E, Andremont A. 2013. Trends in human fecal carriage of extended-spectrum β-lactamases in the community: Toward the globalization of CTX-M. Clin Microbiol Rev 26:744–758.
33. Costelloe C, Metcalfe C, Lovering A, Mant D, Hay AD. 2010. Effect of antibiotic prescribing in primary care on antimicrobial resistance in individual patients: Systematic review and meta-analysis. BMJ 340:1120.
34. Karanika S, Karantanos T, Arvanitis M, Grigoras C, Mylonakis E. 2016. Fecal Colonization with Extended-spectrum Beta-lactamase-Producing Enterobacteriaceae and Risk Factors among Healthy Individuals: A Systematic Review and Metaanalysis. Clin Infect Dis 63:310–318.
35. Woerther PL, Andremont A, Kantele A. 2017. Travel-acquired ESBL-producing Enterobacteriaceae: impact of colonization at individual and community level. J Travel Med 24:S29–S34.
36. Cornejo-Juárez P, Vilar-Compte D, Pérez-Jiménez C, Ñamendys-Silva SA, Sandoval-Hernández S, Volkow-Fernández P. 2015. The impact of hospital-acquired infections with multidrug-resistant bacteria in an oncology intensive care unit. Int J Infect Dis 31:e31–e34.
37. Sabir N, Ikram A, Zaman G, Satti L, Gardezi A, Ahmed A, Ahmed P. 2017. Bacterial biofilm-based catheter-associated urinary tract infections: Causative pathogens and antibiotic resistance. Am J Infect Control.
38. Nguyen DB, Shugart A, Lines C, Shah AB, Edwards J, Pollock D, Sievert D, Patel PR. 2017. National healthcare safety network (NHSN) dialysis event surveillance report for 2014. Clin J Am Soc Nephrol 12:1139–1146.
39. Barlow M. 2009. What Antimicrobial Resistance Has Taught Us About Horizontal Gene Transfer, p. 397–411. *In* Gogarten, MB, Gogarten, JP, Olendzenski, LC (eds.), Horizontal

Gene Transfer: Genomes in Flux. Humana Press, Totowa, NJ.
40. Han JH, Nachamkin I, Tolomeo P, Mao X, Bilker WB, Lautenbach E. 2012. Risk Factors for Efflux Pump Overexpression in Fluoroquinolone-Resistant Escherichia coli. J Infect Dis 206:1597–1603.
41. Nepal K, Pant ND, Neupane B, Belbase A, Baidhya R, Shrestha RK, Lekhak B, Bhatta DR, Jha B. 2017. Extended spectrum beta-lactamase and metallo beta-lactamase production among Escherichia coli and Klebsiella pneumoniae isolated from different clinical samples in a tertiary care hospital in Kathmandu, Nepal. Ann Clin Microbiol Antimicrob 16:1–7.
42. Nordmann P, Naas T, Poirel L. 2011. Global Spread of Carbapenemase- producing Enterobacteriaceae. Emerg Infect Dis 17:1791–1798.
43. Yin W, Li H, Shen Y, Liu Z, Wang S, Shen Z, Zhang R, Walsh TR, Shen J, Wang Y. 2017. Novel Plasmid-Mediated Colistin Resistance Gene mcr-3 in Escherichia coli. MBio 8:00543–17.
44. Peirano G, Bradford PA, Kazmierczak KM, Badal RE, Hackel M, Hoban DJ, Pitout JDD. 2014. Global Incidence of Carbapenemase-Producing Escherichia Coli ST131. Emerg Infect Dis 20:1928–1931.
45. Johnson JR, Johnston B, Clabots C, Kuskowski MA, Castanheira M. 2010. Escherichia coli Sequence Type ST131 as the Major Cause of Serious Multidrug-Resistant E. coli Infections in the United States . Clin Infect Dis 51:286–294.
46. Totsika M, Beatson SA, Sarkar S, Phan MD, Petty NK, Bachmann N, Szubert M, Sidjabat HE, Paterson DL, Upton M, Schembri MA. 2011. Insights into a multidrug resistant Escherichia coli pathogen of the globally disseminated ST131 lineage: Genome analysis and virulence mechanisms. PLoS One 6.
47. Cusumano CK, Hung CS, Chen SL, Hultgren SJ. 2010. Virulence plasmid harbored by uropathogenic Escherichia coli functions in acute stages of pathogenesis. Infect Immun 78:1457–1467.
48. Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. 2005. The microbial pan-genome. Curr Opin Genet Dev 15:589–594.
49. Gillings MR. 2013. Evolutionary consequences of antibiotic use for the resistome, mobilome, and microbial pangenome. Front Microbiol 4:1–10.
50. Vernikos G, Medini D, Riley DR, Tettelin H. 2015. Ten years of pan-genome analyses. Curr Opin Microbiol 23:148–154.
51. Centers for Disease Control and Prevention. 2019. Antibiotic resistance threats in the United States.
52. Colello R, Etcheverría AI, Di Conza JA, Gutkind GO, Padola NL. 2015. Antibiotic

resistance and integrons in shiga toxin-producing escherichia coli (STEC). Brazilian J Microbiol 46:1–5.
53. Raphael E, Wong LK, Riley LW. 2011. Extended-Spectrum Beta-Lactamase Gene Sequences in Gram-Negative Saprophytes on Retail Organic and Nonorganic Spinach. Appl Environ Microbiol 77:1601–1607.
54. Donskey CJ. 2004. The Role of the Intestinal Tract as a Reservoir and Source for Transmission of Nosocomial Pathogens. Clin Infect Dis 39:219–226.
55. Carlet J. 2012. The gut is the epicentre of antibiotic resistance. Antimicrob Resist Infect Control 1:1.
56. Lautenbach E, Patel JB, Bilker WB, Edelstein PH, Fishman NO. 2001. Extended-spectrum β-lactamase-producing Escherichia coli and Klebsiella pneumoniae: Risk factors for infection and impact of resistance on outcomes. Clin Infect Dis 32:1162–1171.
57. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA, Group P-P. 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 4:1.
58. Veritas Health Innovation. Covidence systematic review software. Melbourne, Australia.
59. Van Houwelingen HC, Zwinderman KH, Stijnen T. 1993. A bivariate approach to meta-analysis. Stat Med 12:2273–2284.
60. Mantel N, Haenszel W. 1959. Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease. JNCI J Natl Cancer Inst 22:719–748.
61. Inthout J, Ioannidis JP, Borm GF. 2014. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol 14:1–12.
62. Egger M, Smith GD, Schneider M, Minder C. 1997. Bias in meta-analysis detected by a simple, graphical test. BMJ 315:629–634.
63. R Core Team. 2018. R: A Language and Environment for Statistical Computing. Vienna, Austria.
64. Balduzzi S, Rücker G, Schwarzer G. 2019. How to perform a meta-analysis with {R}: a practical tutorial. Evid Based Ment Health 153–160.
65. Arcilla MS, van Hattem JM, Haverkate MR, Bootsma MCJ, van Genderen PJJ, Goorhuis A, Grobusch MP, Lashof AMO, Molhoek N, Schultsz C, Stobberingh EE, Verbrugh HA, de Jong MD, Melles DC, Penders J. 2017. Import and spread of extended-spectrum beta-lactamase-producing Enterobacteriaceae by international travellers (COMBAT study): a prospective, multicentre cohort study. LANCET Infect Dis 17:78–85.
66. Angelin M, Forsell J, Granlund M, Evengard B, Palmgren H, Johansson A. 2015. Risk factors for colonization with extended-spectrum beta-lactamase producing

Enterobacteriaceae in healthcare students on clinical assignment abroad: A prospective study. Travel Med Infect Dis 13:223–229.

67. Reuland EA, Halaby T, Hays JP, de Jongh DMC, Snetselaar HDR, van Keulen M, Elders PJM, Savelkoul PHM, Vandenbroucke-Grauls CMJE, Al Naiemi N. 2015. Plasmid-mediated AmpC: prevalence in community-acquired isolates in Amsterdam, the Netherlands, and risk factors for carriage. PLoS One 10:e0113033.
68. Ruh E, Zakka J, Hoti K, Fekrat A, Guler E, Gazi U, Erdogmus Z, Suer K. 2019. Extended-spectrum beta-lactamase, plasmid-mediated AmpC beta-lactamase, fluoroquinolone resistance, and decreased susceptibility to carbapenems in Enterobacteriaceae: fecal carriage rates and associated risk factors in the community of Northern Cyprus. Antimicrob Resist Infect Control 8:98.
69. Sanneh B, Kebbeh A, Jallow HS, Camara Y, Mwamakamba LW, Ceesay IF, Barrow E, Sowe FO, Sambou SM, Baldeh I, Jallow A, Raul MAJ, Andremont A. 2018. Prevalence and risk factors for faecal carriage of Extended Spectrum beta-lactamase producing Enterobacteriaceae among food handlers in lower basic schools in West Coast Region of The Gambia. PLoS One 13.
70. Vading M, Kabir MH, Kalin M, Iversen A, Wiklund S, Naucler P, Giske CG. 2016. Frequent acquisition of low-virulence strains of ESBL-producing Escherichia coli in travellers. J Antimicrob Chemother 71:3548–3555.
71. Wielders CCH, van Hoek AHAM, Hengeveld PD, Veenman C, Dierikx CM, Zomer TP, Smit LAM, van der Hoek W, Heederik DJ, de Greeff SC, Maassen CBM, van Duijkeren E. 2017. Extended-spectrum beta-lactamase- and pAmpC-producing Enterobacteriaceae among the general population in a livestock-dense area. Clin Microbiol Infect 23:120.e1-120.e8.
72. Caudell MA, Mair C, Subbiah M, Matthews L, Quinlan RJ, Quinlan MB, Zadoks R, Keyyu J, Call DR. 2018. Identification of risk factors associated with carriage of resistant Escherichia coli in three culturally diverse ethnic groups in Tanzania: a biological and socioeconomic analysis. Lancet Planet Heal 2:e489–e497.
73. Dohmen W, Van Gompel L, Schmitt H, Liakopoulos A, Heres L, Urlings BA, Mevius D, Bonten MJM, Heederik DJJ. 2017. ESBL carriage in pig slaughterhouse workers is associated with occupational exposure. Epidemiol Infect 145:2003–2010.
74. Dohmen W, Schmitt H, Bonten M, Heederik D. 2017. Air exposure as a possible route for ESBL in pig farmers. Environ Res 155:359–364.
75. Lubbert C, Straube L, Stein C, Makarewicz O, Schubert S, Mossner J, Pletz MW, Rodloff AC. 2015. Colonization with extended-spectrum beta-lactamase-producing and carbapenemase-producing Enterobacteriaceae in international travelers returning to

Germany. Int J Med Microbiol 305:148–156.
76. McNulty CAM, Lecky DM, Xu-McCrae L, Nakiboneka-Ssenabulya D, Chung K-T, Nichols T, Thomas HL, Thomas M, Alvarez-Buylla A, Turner K, Shabir S, Manzoor S, Smith S, Crocker L, Hawkey PM. 2018. CTX-M ESBL-producing Enterobacteriaceae: estimated prevalence in adults in England in 2014. J Antimicrob Chemother 73:1368–1388.
77. Barreto Miranda I, Ignatius R, Pfuller R, Friedrich-Janicke B, Steiner F, Paland M, Dieckmann S, Schaufler K, Wieler LH, Guenther S, Mockenhaupt FP. 2016. High carriage rate of ESBL-producing Enterobacteriaceae at presentation and follow-up among travellers with gastrointestinal complaints returning from India and Southeast Asia. J Travel Med 23:tav024.
78. Mo Y, Seah I, Lye PSP, Kee XLJ, Wong KYM, Ko KKK, Ong RT-H, Tambyah PA, Cook AR. 2019. Relating knowledge, attitude and practice of antibiotic use to extended-spectrum beta-lactamase-producing Enterobacteriaceae carriage: results of a cross-sectional community survey. BMJ Open 9:e023859.
79. Reuland EA, Al Naiemi N, Kaiser AM, Heck M, Kluytmans JAJW, Savelkoul PHM, Elders PJM, Vandenbroucke-Grauls CMJE. 2016. Prevalence and risk factors for carriage of ESBL-producing Enterobacteriaceae in Amsterdam. J Antimicrob Chemother 71:1076–1082.
80. Butcher CR, Rubin J, Mussio K, Riley LW. 2019. Risk Factors Associated with Community-Acquired Urinary Tract Infections Caused by Extended-Spectrum $β$-Lactamase-Producing Escherichia coli: a Systematic Review . Curr Epidemiol Reports 6:300–309.
81. Nielsen KL, Stegger M, Kiil K, Godfrey PA, Feldgarden M, Lilje B, Andersen PS, Frimodt-Møller N. 2017. Whole-genome comparison of urinary pathogenic Escherichia coli and faecal isolates of UTI patients and healthy controls. Int J Med Microbiol 307:497–507.
82. Thänert R, Reske KA, Hink T, Wallace MA, Wang B, Schwartz DJ, Seiler S, Cass C, Burnham CA, Dubberke ER, Kwon JH, Dantas G. 2019. Comparative Genomics of Antibiotic-Resistant Uropathogens Implicates Three Routes for Recurrence of Urinary Tract Infections. MBio 10:1–16.
83. Zurfluh K, Nüesch-Inderbinen M, Morach M, Zihler Berner A, Hächler H, Stephan R. 2015. Extended-spectrum-β-lactamase-producing Enterobacteriaceae isolated from vegetables imported from the Dominican Republic, India, Thailand, and Vietnam. Appl Environ Microbiol 81:3115–3120.
84. Jongman M, Korsten L. 2016. Genetic diversity and antibiotic resistance of Escherichia

coli isolates from different leafy green production systems. J Food Prot 79:1846–1853.
85. World Health Organization (WHO). 2020. Global Antimicrobial Resistance and Use Surveillance System (GLASS) Report.
86. World Health Organization (WHO). 2015. Global Action Plan on Antimicrobial Resistance https://doi.org/10.1128/microbe.10.354.1.
87. U.S. Department of Health & Human Services. 2020. NATIONAL ACTION PLAN FOR COMBATING ANTIBIOTIC-RESISTANT BACTERIA 2020-2025.
88. Karanika S, Karantanos T, Arvanitis M, Grigoras C, Mylonakis E. 2016. Fecal Colonization with Extended-spectrum Beta-lactamase-Producing Enterobacteriaceae and Risk Factors among Healthy Individuals: A Systematic Review and Metaanalysis. Clin Infect Dis 63:310–318.
89. Rubin J, Mussio K, Xu Y, Suh J, Riley LW. 2020. Prevalence of antimicrobial resistance genes and integrons in commensal Gram-negative bacteria in a college community. Microb Drug Resist 26:1227–1235.
90. Hu Y, Matsui Y, Riley LW. 2020. Risk factors for fecal carriage of drug-resistant Escherichia coli : a systematic review and meta-analysis. Antimicrob Resist Infect Control 3:1–12.
91. Manges AR, Smith SP, Lau BJ, Nuval CJ, Eisenberg JN, Dietrich PS, Riley LW. 2007. Retail meat consumption and the acquisition of antimicrobial resistant Escherichia coli causing urinary tract infections: a case-control study. Foodborne Pathog Dis 4:419–431.
92. Wagenmakers EJ, Farrell S. 2004. AIC model selection using Akaike weights. Psychon Bull Rev 11:192–196.
93. Tibshirani R. 1996. Regression shrinkage and selection via the Lasso. J R Stat Soc Ser B.
94. Hosmer DW, Lameshow S. 1989. Applied Logistic Regression. New York Wiley.
95. DeLong ER, DeLong DM, Clarke-Pearson DL. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics 44:837–845.
96. Miranda IB, Ignatius R, Pfueller R, Friedrich-Jaenicke B, Steiner F, Paland M, Dieckmann S, Schaufler K, Wieler LH, Guenther S, Mockenhaupt FP. 2016. High carriage rate of ESBL-producing Enterobacteriaceae at presentation and follow-up among travellers with gastrointestinal complaints returning from India and Southeast Asia. J Travel Med 23.
97. Yamaji R, Friedman CR, Rubin J, Suh J, Thys E, McDermott P, Hung-Fan M, Riley LW. 2018. A Population-Based Surveillance Study of Shared Genotypes of Escherichia coli Isolates from Retail Meat and Suspected Cases of Urinary Tract Infections. mSphere 3:e00179-18.
98. Liao N, Borges CA, Rubin J, Hu Y, Ramirez HA, Chen J, Zhou B, Zhang Y, Zhang R,

Jiang J, Riley LW. 2020. Prevalence of β-Lactam Drug-Resistance Genes in Escherichia coli Contaminating Ready-to-Eat Lettuce. Foodborne Pathog Dis 17:739–742.
99. Tyson GH, Li C, Hsu CH, Bodeis-Jones S, McDermott PF. 2019. Diverse Fluoroquinolone Resistance Plasmids From Retail Meat E. coli in the United States. Front Microbiol 10:1–7.
100. Ryu SH, Park SG, Choi SM, Hwang YO, Ham HJ, Kim SU, Lee YK, Kim MS, Park GY, Kim KS, Chae YZ. 2012. Antimicrobial resistance and resistance genes in Escherichia coli strains isolated from commercial fish and seafood. Int J Food Microbiol 152:14–18.
101. Buschmann AH, Tomova A, López A, Maldonado MA, Henríquez LA, Ivanova L, Moy F, Godfrey HP, Cabello FC. 2012. Salmon aquaculture and antimicrobial resistance in the marine environment. PLoS One 7:26–28.
102. Søraas A, Sundsfjord A, Sandven I, Brunborg C, Jenum PA. 2013. Risk Factors for Community-Acquired Urinary Tract Infections Caused by ESBL-Producing Enterobacteriaceae -A Case-Control Study in a Low Prevalence Country. PLoS One 8:1–7.
103. Schmedes M, Brejnrod AD, Aadland EK, Kiilerich P, Kristiansen K, Jacques H, Lavigne C, Graff IE, Eng Ø, Holthe A, Mellgren G, Young JF, Sundekilde UK, Liaset B, Bertram HC. 2019. The Effect of Lean-Seafood and Non-Seafood Diets on Fecal Metabolites and Gut Microbiome: Results from a Randomized Crossover Intervention Study. Mol Nutr Food Res 63:1–8.
104. Bratlie M, Hagen I V., Helland A, Erchinger F, Midttun Ø, Ueland PM, Rosenlund G, Sveier H, Mellgren G, Hausken T, Gudbrandsen OA. 2020. Effects of high intake of cod or salmon on gut microbiota profile, faecal output and serum concentrations of lipids and bile acids in overweight adults: a randomised clinical trial. Eur J Nutr https://doi.org/10.1007/s00394-020-02417-8.
105. Yassin AK, Gong J, Kelly P, Lu G, Guardabassi L, Wei L, Han X, Qiu H, Price S, Cheng D, Wang C. 2017. Antimicrobial resistance in clinical Escherichia coli isolates from poultry and livestock, China. PLoS One 12:1–8.
106. Borges CA, Tarlton NJ, Riley LW. 2019. Escherichia coli from Commercial Broiler and Backyard Chickens Share Sequence Types, Antimicrobial Resistance Profiles, and Resistance Genes with Human Extraintestinal Pathogenic Escherichia coli. Foodborne Pathog Dis 16:813–822.
107. Y. M, T. N, M. T, T. K, M. Y, M. N, S. T, S. I, O. Y, T. H, M. S, H. W, T. S, A. H, T. H, N. F, T. K, Y. Y, K. U, G. Y, K. M, S. Y. 2017. Population structure of Japanese extraintestinal pathogenic Escherichia coli and its relationship with antimicrobial resistance. J Antimicrob Chemother 72:1040–1049.

108. Yamaji R, Rubin J, Thys E, Friedman CR, Riley LW. 2018. Persistent pandemic lineages of uropathogenic Escherichia coli in a college community from 1999 to 2017. J Clin Microbiol 56:1–10.
109. Li D, Reid CJ, Kudinha T, Jarocki VM, Djordjevic SP. 2020. Genomic analysis of trimethoprim-resistant extraintestinal pathogenic Escherichia coli and recurrent urinary tract infections. Microb Genomics 6:1–19.
110. Ludden C, Decano AG, Jamrozy D, Pickard D, Morris D, Parkhill J, Peacock SJ, Cormican M, Downing T. 2020. Genomic surveillance of escherichia coli ST131 identifies local expansion and serial replacement of subclones. Microb Genomics 6.
111. Kallonen T, Brodrick HJ, Harris SR, Corander J, Brown NM, Martin V, Peacock SJ, Parkhill J. 2017. Systematic longitudinal survey of invasive Escherichia coli in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. Genome Res 27:1437–1449.
112. McNally A, Kallonen T, Connor C, Abudahab K, Aanensen DM, Horner C, Peacock SJ, Parkhill J, Croucher NJ, Corander J. 2019. Diversification of colonization factors in a multidrug-resistant escherichia coli lineage evolving under negative frequency- dependent selection. MBio 10:1–19.
113. Adams-Sapper S, Diep BA, Perdreau-Remington F, Riley LW. 2013. Clonal composition and community clustering of drug-susceptible and -resistant Escherichia coli isolates from bloodstream infections. Antimicrob Agents Chemother 57:490–497.
114. Tartof SY, Solberg OD, Manges AR, Riley LW. 2005. Analysis of a Uropathogenic Escherichia coli Clonal Group by Multilocus Sequence Typing. J Clin Microbiol 43:5860–5864.
115. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data.
116. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:1–22.
117. Zhou Z, Alikhan NF, Mohamed K, Fan Y, Achtman M. 2020. The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity. Genome Res 30:138–152.
118. Seemann T. 2014. ABRicate.
119. Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.
120. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: Architecture and applications. BMC Bioinformatics 10:1–9.
121. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: Rapid large-scale prokaryote pan genome analysis.

Bioinformatics 31:3691–3693.
122. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR. 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb genomics 2:e000056.
123. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One 5.
124. Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
125. Stephens C, Adams-Sapper S, Sekhon M, Johnson J, Riley L. Genomic analysis of Escherichia coli ST95 isolates from bloodstream infections reveals multiple sub-lineages with distinct plasmids and antibiotic resistance gene content. https://doi.org/10.1128/EC.4.3.604.
126. Dias RCS, Moreira BM, Riley LW. 2010. Use of fimH single-nucleotide polymorphisms for strain typing of clinical isolates of Escherichia coli for epidemiologic investigation. J Clin Microbiol 48:483–488.
127. Jeckelmann JM, Erni B. 2020. Transporters of glucose and other carbohydrates in bacteria. Pflugers Arch Eur J Physiol 472:1129–1153.
128. Garcia De Gonzalo C V., Denham EL, Mars RAT, Stülke J, Van Der Donk WA, Van Dijl JM. 2015. The phosphoenolpyruvate: Sugar phosphotransferase system is involved in sensitivity to the glucosylated bacteriocin sublancin. Antimicrob Agents Chemother 59:6844–6854.
129. Mann R, Mediati DG, Duggin IG, Harry EJ, Bottomley AL. 2017. Metabolic adaptations of Uropathogenic E. coli in the urinary tract. Front Cell Infect Microbiol 7.
130. Su M, Satola SW, Read TD. 2019. Genome-based prediction of bacterial antibiotic resistance. J Clin Microbiol 57.
131. Zankari E, Hasman H, Kaas RS, Seyfarth AM, Agersø Y, Lund O, Larsen MV, Aarestrup FM. 2013. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. J Antimicrob Chemother 68:771–777.