Contents lists available at ScienceDirect

One Health

journal homepage: www.elsevier.com/locate/onehlt

Using source-associated mobile genetic elements to identify zoonotic extraintestinal *E. coli* infections

Cindy M. Liu [a,b], Maliha Aziz [a], Daniel E. Park [a], Zhenke Wu [c,d], Marc Stegger [e], Mengbing Li [c], Yashan Wang [a], Kara Schmidlin [f], Timothy J. Johnson [g], Benjamin J. Koch [h], Bruce A. Hungate [h], Lora Nordstrom [f], Lori Gauld [i], Brett Weaver [f], Diana Rolland [i], Sally Statham [f], Brantley Hall [j], Sanjeev Sariya [a], Gregg S. Davis [a], Paul S. Keim [b,f], James R. Johnson [k], Lance B. Price [a,f,*]

[a] *Antibiotic Resistance Action Center, Department of Environmental and Occupational Health, Milken Institute School of Public Health, George Washington University, 800 22nd Street NW, Washington, DC 20052, USA*
[b] *The Pathogen and Microbiome Institute, Department of Biological Sciences, Northern Arizona University, Room 210 Building 56, Applied Research & Development, 1395 S Knoles Drive, Flagstaff, AZ 86011, USA*
[c] *Department of Biostatistics, University of Michigan School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109, USA*
[d] *Michigan Institute for Data Science (MIDAS), University of Michigan, 500 Church Street, Suite 600, Ann Arbor, MI 48109, USA*
[e] *Department of Bacteria, Parasites and Fungi, Statens Serum Institut, 5 Artillerivej, DK-2300 Copenhagen, Denmark*
[f] *Division of Pathogen Genomics, Translational Genomics Research Institute (TGen), 3051 W Shamrell Blvd, Flagstaff, AZ 86005, USA*
[g] *Department of Veterinary and Biomedical Sciences, College of Veterinary Medicine, University of Minnesota, 1365 Gortner Ave, St Paul, MN 55108, USA*
[h] *Center for Ecosystem Science and Society, Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA*
[i] *Flagstaff Medical Center, 1200 N. Beaver St. Flagstaff, AZ 86001, USA*
[j] *Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742, USA*
[k] *Minneapolis Veterans Affairs Health Care System, 1 Veterans Dr, Minneapolis, MN 55417, USA*

ARTICLE INFO

Keywords:
Escherichia coli
Infectious diseases
Urinary tract infection
Zoonosis
Foodborne
Mobile genetic element
Host adaptation
Food-animal production
Bayesian latent class model
Public health

ABSTRACT

A one-health perspective may provide new and actionable information about *Escherichia coli* transmission. *E. coli* colonizes a broad range of vertebrates, including humans and food-production animals, and is a leading cause of bladder, kidney, and bloodstream infections in humans. Substantial evidence supports foodborne transmission of pathogenic *E. coli* strains from food animals to humans. However, the relative contribution of foodborne zoonotic *E. coli* (FZEC) to the human extraintestinal disease burden and the distinguishing characteristics of such strains remain undefined. Using a comparative genomic analysis of a large collection of contemporaneous, geographically-matched clinical and meat-source *E. coli* isolates ($n = 3111$), we identified 17 source-associated mobile genetic elements – predominantly plasmids and bacteriophages – and integrated them into a novel Bayesian latent class model to predict the origins of clinical *E. coli* isolates. We estimated that approximately 8 % of human extraintestinal *E. coli* infections (mostly urinary tract infections) in our study population were caused by FZEC. FZEC strains were equally likely to cause symptomatic disease as non-FZEC strains. Two FZEC lineages, ST131-*H*22 and ST58, appeared to have particularly high virulence potential. Our findings imply that FZEC strains collectively cause more urinary tract infections than does any single non-*E. coli* uropathogenic species (e.g., *Klebsiella pneumoniae*). Our novel approach can be applied in other settings to identify the highest-risk FZEC strains, determine their sources, and inform new one-health strategies to decrease the heavy public health burden imposed by extraintestinal *E. coli* infections.

Abbreviations: BLCM, Bayesian latent class model; FZEC, Foodborne zoonotic *Escherichia coli*; MGE, Mobile genetic element; MLST, multilocus sequence typing; ST, Sequence type; UTI, urinary tract infections.

* Corresponding author at: Antibiotic Resistance Action Center, Milken Institute School of Public Health, George Washington University, Science & Engineering Hall, 800 22nd Street NW, 7th Floor - Office 7590, Washington, DC 20052, USA.

E-mail address: lprice@gwu.edu (L.B. Price).

https://doi.org/10.1016/j.onehlt.2023.100518
Received 30 September 2022; Received in revised form 23 February 2023; Accepted 24 February 2023
Available online 28 February 2023

1. Background

The concept that urinary tract infections (UTIs) could be caused by foodborne *Escherichia coli* was proposed as early as the 1960s as an explanation for both community- and hospital-acquired UTI [1–3]. Subsequent investigations supported this concept for both UTI outbreaks [4,5] and sporadic cases [6,7]. Although *E. coli* UTI only rarely progresses to serious invasive disease, the enormous number of cases makes it a leading cause of *E. coli*-associated sepsis deaths worldwide [8]. Yet, we still lack a fundamental understanding of the origins and transmission dynamics of the causative *E. coli* strains, including the proportion that originate from food animals and are acquired through the food supply (i.e., foodborne zoonotic *E. coli* (FZEC)). Until we close these knowledge gaps, our ability to control these prevalent infections will be limited.

Most UTIs are sporadic; therefore, quantifying the overall proportion of cases that are attributable to foodborne zoonotic *E. coli* strains differs fundamentally from identifying the source of single-strain foodborne disease outbreaks [9,10]. That is because the extraintestinal pathogenic *E. coli* strains that cause UTIs, colonize the billions of animals raised for meat, and contaminate meat products are tremendously diverse [11]. Each *E. coli* sequence type (ST) comprises multiple distinct genetic variants (strains), and the total *E. coli* population currently comprises >13,000 STs [12–15]. Thus, identifying direct strain matches between UTI isolates and meat-source isolates is extremely unlikely, even with a huge sample size. Accordingly, defining the fraction of UTIs attributable to FZEC strains requires a novel genomic approach.

Host-adaptive genes, which likely underlie the broad host range of *E. coli*, may hold the key to differentiating *E. coli* strains that originate from various animal species [16]. As much as 47% of the genes in an individual *E. coli* strain can be categorized as “accessory genes” and may help it adapt to specific environments [17]. The host-adaptive genes that are on mobile genetic elements (MGEs), in particular, may be lost and gained quickly as *E. coli* strains transition between hosts. *E. coli* carrying MGEs inconsistent with the vertebrate species from which they were isolated may be indicative of recent host transitions (i.e., spillovers). Previously, we used poultry-associated ColV plasmids in conjunction with high-resolution phylogenetics to identify poultry-adapted *E. coli* strains among human UTI isolates [18]. ColV plasmids and similar host-associated MGEs could serve as powerful tools for quantifying the proportion of human extraintestinal infections caused by FZEC [19].

In the present study, we compared the genomes of *E. coli* from retail meat products with contemporaneous human clinical isolates from the same locale to identify source-associated MGEs. By using a novel Bayesian statistical approach that incorporated these elements, we further inferred the clinical isolates’ likely host origin, then estimated the proportion of human *E. coli* cystitis episodes or invasive infections caused by foodborne zoonotic *E. coli*. We also sought to identify meat-source *E. coli* lineages that are disproportionately likely to cause foodborne cystitis or invasive infections in humans.

2. Methods

2.1. Study population

Retail meat samples and clinical isolates were collected from January 1, 2012 to December 31, 2012 in Flagstaff, Arizona as described previously [18]. During this 12-month period, all available brands of raw chicken, turkey, and pork were sampled from all nine major grocery chains in Flagstaff every two weeks. Concurrently, all human clinical *E. coli* isolates from urine and blood samples at the Flagstaff Medical Center – the main clinical laboratory serving Flagstaff and surrounding cities – were collected. The Northern Arizona Healthcare IRB approved isolate collection and medical record review (protocol number: 573857–4) with a waiver of consent.

2.2. *E. coli* collection and susceptibility testing

A single *E. coli* isolate was recovered from each meat product using enrichment methods as described previously [20]. Antimicrobial susceptibility for the meat isolates was determined by disk diffusion in accordance with the Clinical and Laboratory Standards Institute [21].

Urine and blood *E. coli* isolates were recovered in the clinical laboratory using standard techniques. Briefly, urine specimens were collected by midstream clean-catch or straight catheterization and cultured on sheep blood and MacConkey agars within 2 h of collection, or with refrigeration for up to 24 h. A positive urine culture was defined as $\geq 10^4$ colony-forming units (CFU)/mL of urine for clean-catch or $\geq 10^3$ CFU/mL for straight catheterization specimens. Species determination and antimicrobial susceptibility testing was performed using the BD Phoenix (Becton-Dickinson Diagnostic Systems, Sparks, MD, USA).

2.3. Clinical data collection

For patients with positive blood and urine *E. coli* cultures, retrospective medical record review was performed whenever possible to collect primary and secondary diagnoses, clinical laboratory results, and symptom keywords. The clinical syndrome associated with each isolate was annotated as asymptomatic bacteriuria, cystitis, pyelonephritis, bacteremia/sepsis, or urosepsis (Figs. S1 and S2). Blood and urine isolates collected during a 48 h period from patients with identical age, sex, and city of residence were annotated as being from a single urosepsis case ($n = 25$ pairs); however, both blood and urine isolates were included in subsequent genomic analyses.

2.4. Core-genome-based phylogenetic analysis

Illumina short-read DNA sequence libraries were generated from 1,188 human clinical urine and blood isolates and 1,923 meat isolates, including chicken ($n = 1{,}156$), turkey ($n = 473$), and pork samples ($n = 310$). Genomes were assembled and characterized by multilocus sequence typing (MLST) as described previously [18]. Pan-MLST and individual MLST maximum-likelihood phylogenies were generated as described previously [18]. A list of reference genomes is available in the supplementary materials (Table S1).

2.5. Identification of source-associated accessory genes

Illumina DNA reads were assembled and annotated with Prokka (v.1.13) and the resultant GFF files were used in Roary (v.3.12.0) for pan-genome analysis to generate accessory gene presence/absence matrix for each isolate. Accessory genes differentially associated with human, chicken, turkey, and pork sample types were identified using pairwise comparisons (i.e., human vs. an individual meat type) based on sensitivity (> 25%), specificity (> 60%), and odds ratio (> 1), Benjamini Hochberg *p*-value (< 0.05) and best pairwise p-value (< 0.05). Source-associated accessory genes were further collapsed into gene clusters (i.e., elements) when their pairwise Pearson correlation was >0.7.

2.6. Inferring host origin by Bayesian latent class model

A Bayesian latent class model (BLCM) was used to generate probabilistic predictions of host origin for each isolate as previously described [22]. The BLCM applied here assumes the host origin for each isolate is in an unobserved class of human or meat. To infer the latent host origins, BLCM uses multivariate binary responses from presence or absence of 17 source-associated MGEs, along with clade information based on the pan-ST core genome phylogeny, to generate a host-origin probability score for each isolate. For response probabilities, independent logistic normal priors with mean 0 and standard deviation 1.5 were used; for the class probabilities (human vs meat), a Beta (1,1) prior was used.

2.7. *Statistical analysis*

Demographic, antimicrobial susceptibility, and clinical variables were compared between putative zoonotic isolates and those of putative human origin using Chi-squared, Fisher's Exact, and Wilcoxon Rank Sum tests, as appropriate. The association between age, sex and putative origin category was evaluated using multivariate logistic regression. Missing data were excluded from statistical comparisons. Analyses were performed in SAS (SAS Institute Inc., v.9.4, Cary, NC, USA), R, v.3.3.1, and with Bayesian inference software JAGS v.4.2.0.

2.8. *Role of the funding source*

The funders had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

3. Results

3.1. *Core genome diversity among E. coli isolates from retail meat and human infections*

Significant population diversity was evident among our contemporaneously collected retail meat and human clinical *E. coli* isolates in the study site (Flagstaff, Arizona) over a 12-month period. Some sequence types (STs) consisted solely of meat isolates or human isolates; others included *E. coli* isolates from both sample types. Specifically, we detected 443 STs among the 3120 total *E. coli* isolates (1932 retail meat isolates; 1188 human clinical blood and urine isolates). Of the 443 STs, 247 included only meat isolates, 120 included only human isolates, and 76 included both meat and human isolates.

We generated a pan-ST phylogenetic tree, which revealed deep-rooted relationships between lineage and sample type (meat versus human), while also showing evidence of host transitions (Fig. 1). The pan-ST tree included three major clades: clades CH1 and CH2 comprised mainly human isolate-dominated STs, whereas clade CM1 comprised mainly meat isolate-dominated STs. This extensive phylogenetic segregation of meat and human isolates suggests host-specific biological adaptation. Given that background, the intermixing of human and meat-source isolates within certain clades and individual STs suggests host transitions.

To investigate potential zoonotic *E. coli* transmissions, we generated rooted phylogenies for the 56 individual STs that included four or more *E. coli* isolates from meat and humans (Fig. S3). In these higher-resolution rooted phylogenies, we expected to accomplish two goals: first, to identify the direction of transmission based on the locations of meat and human isolates in ancestral versus derived clades; second, to infer the time scale of transmission events based on short or long branch lengths found in each clade. However, the resulting phylogenies revealed substantial genetic diversity of the *E. coli* populations, with few close clonal relationships even among isolates from the same ST and sample type (Fig. S4). This suggested that core-genome phylogenetic analysis alone could not readily be used to identify recent zoonotic transmission events. We therefore probed the accessory genome to gather more information regarding potential host transitions.

3.2. *Identifying E. coli MGEs differentially associated with E. coli from humans versus meat*

Our goal was to identify MGEs that are differentially associated with *E. coli* strains from humans and three major meat types, including chicken, turkey, and pork. To identify such source-associated MGEs, we first constructed a pan-genome using our isolate collection. This pan-genome comprised 60,648 genes, including 2940 core genes (5%) and 57,474 accessory genes (95%). Next, we conducted a genome-wide association study to identify accessory genes that were associated with specific sample types. This revealed 366 non-redundant source-associated accessory genes. Most of these accessory genes (96%) were part of multigene clusters ($n = 46$), of which 17 (37%) bore hallmarks of MGEs (Fig. S5).

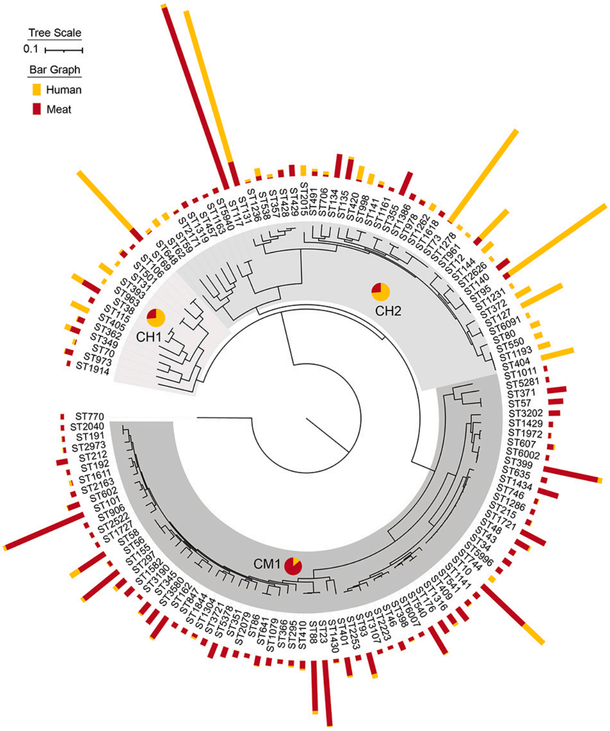

Fig. 1. Phylogenetic relationships and source associations of major sequence types (STs). A rooted, maximum-likelihood phylogeny was constructed using a single arbitrarily-chosen isolate from each ST observed four or more times. The three major clades – CM1, CH1, and CH2 – are shaded (inner ring). For each clade, the proportion of meat and human clinical isolates is shown in a pie chart (red area = meat isolates; yellow area = human clinical isolates). The radiating colored columns indicate the number of isolates in the corresponding ST (proportional to column length, ranging from a minimum of four to a maximum of 223 [ST117]) and their source distribution (reflected by the colour-coding: red area = meat isolates; yellow area = human clinical isolates). Scale bar represents the number of nucleotide substitutions per site. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The 17 source-associated MGEs included 11 plasmids, four prophages, one integrative and conjugative element, and one integrative and mobilizable element. Each MGE contained between 2 and 42 source-associated accessory genes, which co-occurred consistently across strains. Of the 17 source-associated MGEs, six were associated with humans and the other 11 with meat; their known features and possible selective functions are shown Table S2.

We observed that the meat-associated MGEs provided the greatest predictive value for distinguishing between meat versus human isolates, as opposed to distinguishing individual meat types (Fig. 2). Using the 11 meat-associated MGEs collectively – that is, by attributing isolates to a meat source based on presence of at least one of these 11 MGEs – increased the sensitivity for detecting meat isolates to 98.2% from the 70.1% maximum sensitivity of any single meat-associated MGE; however, this approach also lowered the specificity to 34.9% (Fig. 2). The human-associated MGEs could also be used collectively, but again with reduced specificity. To optimize the combined sensitivity and specificity of the 17 MGEs for predicting the origins of *E. coli* isolates, we integrated them into a Bayesian latent class model (BLCM) [22].

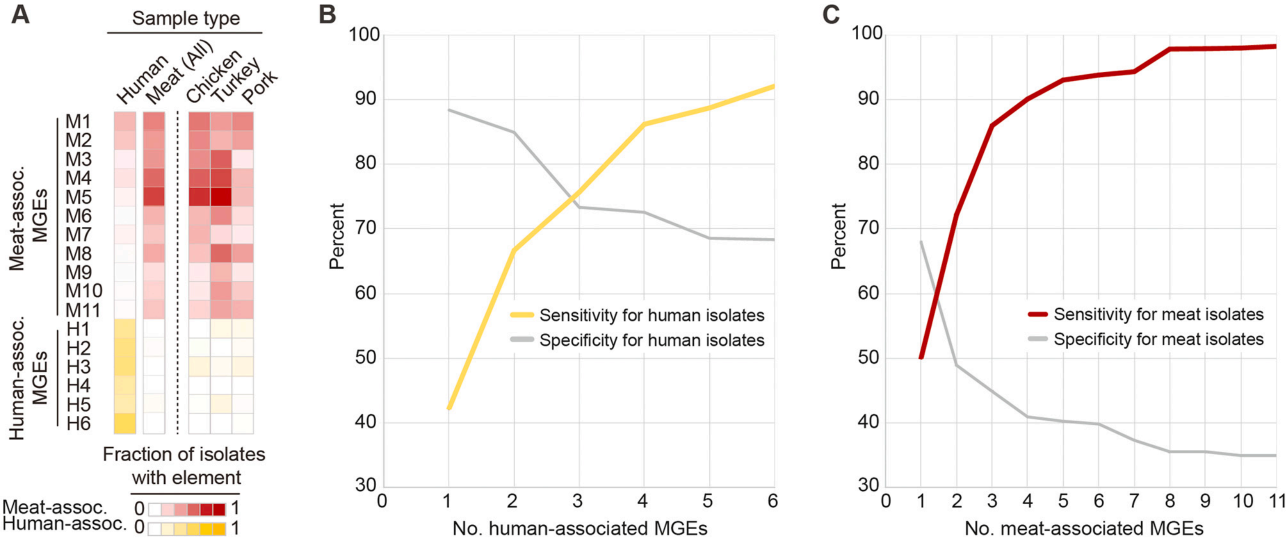

Fig. 2. Prevalence of source-associated mobile genetic elements (MGEs) by sample type and cumulative sensitivity/specificity. Panel A: The prevalence of MGEs among isolates from each sample type, with darker shading indicating higher prevalence of the MGE. Meat-associated MGEs were frequently associated with more than one type of meat (chicken, turkey, pork). Panel B: Sensitivity versus specificity of human-associated MGEs for the detection of *E. coli* strains isolated from human clinical samples, including asymptomatic bacteriuria, cystitis and sepsis. Panel C: Sensitivity versus specificity of meat-associated MGEs for the detection of *E. coli* strains isolated from meat.

3.3. Evaluation and application of a Bayesian latent class model for identifying foodborne zoonotic E. coli

In general, BLCM is a finite mixture model for multivariate discrete responses that can be used to estimate the membership of each data point in a prespecified number of unobserved groups. Here, we used a two-class BLCM for the MGE data to presumptively identify human clinical *E. coli* isolates acquired from food animals through contaminated meat (i.e., foodborne zoonotic *E. coli* [FZEC]). The model was validated on a curated subset of the present isolates, then applied to the entire collection, as described below:

The BLCM assumed that the origin for each isolate was in a latent (i.e., unobserved) class of either meat or human. To infer the latent class of each isolate, the model used placement within one of the three major clades of the pan-ST phylogeny and presence (or absence) of the 17 source-associated MGEs to generate a final latent class posterior probability score for each isolate. To minimize the risk of misclassification during the validation process, we excluded isolates ($n = 168$) that appeared to have undergone a host transition based on phylogenetic analysis.

Our evaluation criterion for the model was agreement between the model's predicted origin (meat or human) and the actual sample type (anticipating some level of disagreement due to the central hypothesis that spillover events do occur). We randomly partitioned the curated isolate collection into two datasets, one containing two-thirds of the data (the training set), the other containing one-third of the data (the validation set). With the validation-set isolates, 96.3% ($n = 949/985$) of the model's predicted source agreed with the known sample type. Model stability assessments, which were done using Markov chain Monte Carlo simulation trace plots and convergence diagnostics, supported the robustness of the BLCM for origin predictions. Finally, we applied the model to the entire dataset.

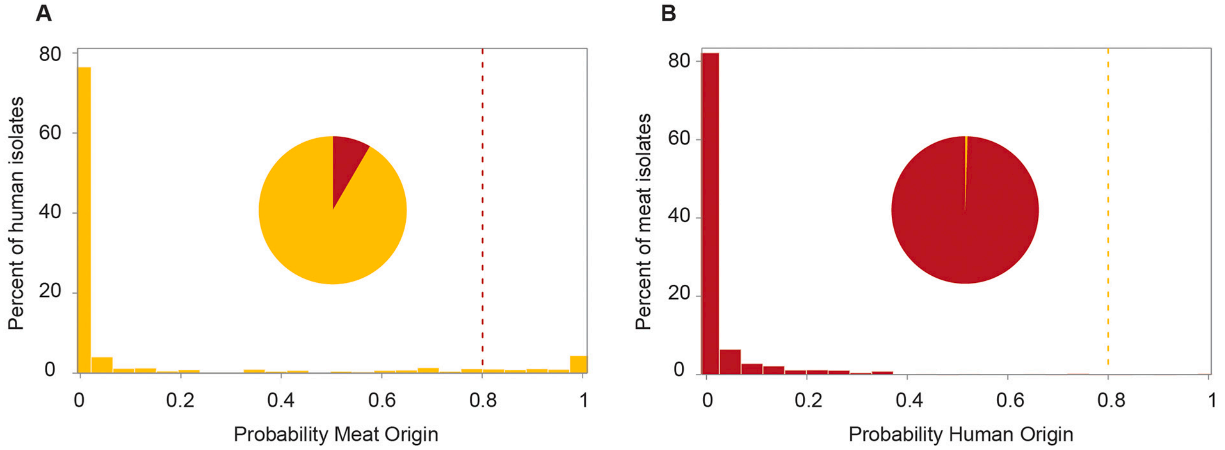

Fig. 3. Bayesian latent class model (BLCM) predictions versus known sample type. Panel A: BLCM probability of meat origin for *E. coli* isolates from human clinical samples. Isolates with probabilities >= 0.8 (right of the red dotted line) were considered to be of meat origin (i.e., meat-to-human spillover) and categorized as foodborne zoonotic *E. coli*. Pie chart shows the proportion of isolates predicted to be of human-origin (yellow) and meat-origin (red). Panel B: BLCM probability of meat origin for isolates from meat samples. Isolates with probabilities >= 0.8 (right of the yellow dotted line) were considered to be of human origin (i.e., human-to-meat spillover). Pie chart shows the proportion of isolates predicted to be of human-origin (yellow) and meat-origin (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The resulting BLCM origin predictions for the full isolate collection showed a strikingly bimodal distribution that was largely consistent with sample type (Fig. 3). To obtain a dichotomous classification (i.e., meat or human origin), we applied user-defined probability thresholds of ≥0.80 for positive and ≤ 0.20 for negative; probabilities between these two values were considered indeterminate.

Detailed analyses comparing inferences from core genome phylogenies versus BLCM are shown in the supplementary materials (Fig. S6 and S7).

3.4. Identification of foodborne zoonotic E. coli strains among human E. coli infections

The BLCM predicted that only a small minority of isolates had a different origin than the known sample type, presumptively indicating a recent host transition (n = 108) (Table S3). Specifically, the model predicted that only 0.5% (n = 10) of meat isolates originated in humans. In contrast, among the 1162 unique clinical cases, the model identified 8.4% (n = 98) as involving *E. coli* of meat origin. Hereafter, we regard the isolates from these 98 human cases as foodborne zoonotic *E. coli* (FZEC).

3.5. Differences in antimicrobial susceptibility profiles support foodborne zoonotic E. coli as a distinct population

Our data indicate that the FZEC population, although presumably derived from meat and recovered from human patients, has a distinct antimicrobial susceptibility profile, as compared with the meat and human-origin *E. coli* populations for which the BLCM origin predictions matched sample type (Table 1). The FZEC isolates were generally less likely to be resistant to antimicrobials than were either the meat or the human-origin isolates. Clinically-relevant exceptions included trimethoprim-sulfamethoxazole and ciprofloxacin, for which resistance was more frequent among FZEC isolates than among meat isolates, and cefazolin and ceftriaxone, for which resistance was similarly frequent among FZEC isolates and human-origin isolates. Thus, although these FZEC isolates were generally less concerning from a resistance perspective than either of the other two populations, they nonetheless exhibited a substantial prevalence of resistance to clinically relevant antimicrobials, including those commonly used to treat *E. coli* infections such as cystitis, pyelonephritis, and sepsis.

Table 1

Antimicrobial susceptibility of meat isolates, foodborne zoonotic *E. coli* (FZEC) clinical isolates, and human-origin clinical isolates.

	Meat Isolates	**Human Clinical Isolates (n = 1188)**			
	A. Meat origin[a] (n = 1827)[d]	**B. FZEC[b] (n = 98)**	**C. non-FZEC[c] (n = 1090)**	**Chi-square p-value[e]**	
	% Isolates (n)			**A vs B**	**B vs C**
Resistance to Antimicrobial Classes					
0	32.5 (594)	51.0 (50)	44.6 (486)		
≥ 1	67.5 (1233)	49.0 (48)	55.4 (604)	**<0.001**	0.22
≥ 2	44.1 (807)	34.7 (34)	49.6 (541)	**0.002**	**0.03**
≥ 3 (multidrug resistant)	29.1 (531)	24.5 (24)	38.3 (417)	**0.01**	**0.02**
Resistance to Individual Antimicrobials					
Ampicillin	33.0 (603)	27.6 (27)	46.7 (509)	0.26	**<0.001**
Ampicillin Sulbactam	24.1 (441)	23.5 (23)	43.0 (469)	0.88	**<0.001**
Cefazolin	29.4 (537)	15.3 (15)	18.3 (199)	**0.003**	0.46
Cefoxitin	12.0 (219)	9.2 (9)	6.2 (68)	0.40	0.26
Ceftriaxone	10.1 (185)	2.0 (2)	2.8 (31)	**0.009**	0.64
Ciprofloxacin	0.1 (1)	6.1 (6)	18.7 (204)	**<0.001**	**0.002**
Gentamicin	24.2 (442)	6.1 (6)	5.0 (54)	**<0.001**	0.61
Tetracycline	49.9 (912)	29.6 (29)	24.3 (265)	**<0.001**	0.25
Trimethoprim Sulfamethoxazole	5.6 (103)	11.2 (11)	28.0 (305)	**0.02**	**<0.001**

Bold p-values denote statistical significance at the $p < 0.05$ level using a Chi-squared test.

[a] Meat-derived isolates identified as having greater than 80% probability of being from a meat source.

[b] Human clinical isolates identified as putative meat-to-human spillover, defined as having greater than 80% probability of being from a meat source.

[c] Human clinical isolates not identified as putative meat-to-human spillover, defined as having less than 80% probability of being from a meat source.

[d] Susceptibility not available for 4 isolates.

[e] Bold values denote statistical significance at the $p < 0.05$ level.

3.6. FZEC were as likely to cause sepsis, cystitis, and asymptomatic bacteriuria as were human-origin E. coli

The clinical syndromes described among the 1162 clinical cases included asymptomatic bacteriuria (41.2%), cystitis (25.2%), pyelonephritis (9.2%) and sepsis/urosepsis (8.6%). Diagnoses were unavailable for 183 (15.7%) cases.

The only difference in clinical syndrome distribution between FZEC-associated cases and human-origin cases was that pyelonephritis was more common among human-origin cases (9.9%, vs. 2.0%: $P = 0.01$) (Table 2). Notably, FZEC-associated cases were as likely to qualify as sepsis or cystitis as putative human-origin cases.

Most patients were female (88.2%). The population's median age was 50 years (IQR, 24–71 years). Male and female patients were affected equally by FZEC and human-origin *E. coli* (Table 2); however, female patients affected by FZEC were significantly older than the female patients affected by human-origin *E. coli.*

3.7. Assessing the zoonotic potential of meat-associated lineages

We next assessed which STs are highest-risk for causing symptomatic or invasive human infections (Fig. 4, Table S4). Multiple STs exhibited a substantial difference in proportional abundance among FZEC isolates vs. among meat isolates. This suggests that intrinsic zoonotic potential varies by ST, such that an ST's abundance among clinical isolates is not merely a stochastic phenomenon that reflects the ST's abundance in meat. Similarly, among the FZEC isolates, the rank order of STs by absolute abundance varied by clinical context, evidence suggesting ST-specific differences in pathogenic potential (Fig. 4, Table S4). For example, ST58 and ST131 were associated with the same fraction of total FZEC cases, which included a large proportion of isolates associated with asymptomatic bacteriuria (ASB). However, when the analysis was constrained to symptomatic infection, ST131 became the dominant FZEC lineage. Yet when the analysis was constrained further to only invasive disease, ST58 became the dominant FZEC lineage, accounting for two cases, as compared to only a single case each for the other eight STs. Additionally, despite being the dominant symptomatic infection FZEC lineages, ST131 and ST58 were relatively infrequent among meat isolates (< 2% each) – further evidence that these two lineages have a relatively high intrinsic virulence potential.

Table 2
Summary of clinical outcomes and patient profiles (age/sex) of foodborne zoonotic *E. coli* (FZEC) vs non-FZEC infections.

	A. FZECclinical[a]	B. Non-FZECclinical[b]	*p*-value[c]
	(n = 98)	(*n* = 969)	AvsB
Diagnosis (n, column %)			
Asymptomatic bacteriuria	44 (44.9)	386 (39.8)	0.33
Cystitis	32 (32.7)	240 (24.8)	0.09
Pyelonephritis	2 (2.0)	96 (9.9)	**0.01**
Sepsis/urosepsis	8 (8.2)	86 (8.9)	0.81
Not coded	12 (12.2)	161 (16.6)	0.27
Sex (n, column %)			
Female	91 (92.9)	847 (87.4)	0.12
Male	7 (7.1)	122 (12.6)	
Age (median, *Q1-Q3*)	63 (35–78)	50 (24–71)	0.004
Female	63 (38–78)	47 (24–70)	**0.001**
Male	46 (28–73)	63 (46–75)	0.34

[a] Human clinical isolates identified as putative meat-to-human spillover, defined as having greater than 80% probability of being from a meat source.

[b] Human clinical isolates identified as unlikely to be a putative meat-to-human spillover, defined as having less than 20% probability of being from a meat source.

[c] Bold values denote statistical significance at the Chi-square $p < 0.05$ level.

4. Discussion

In this study, we sought to develop new tools for studying *E. coli* through a one-health lens and estimate the proportion of human extraintestinal *E. coli* infections caused by foodborne zoonotic *E. coli* (FZEC) strains. We accomplished this by identifying 17 source-associated mobile genetic elements (MGEs) and incorporating them into a Bayesian latent class model to predict the origin of *E. coli* isolates. Using this approach, we determined that approximately 8% of human extraintestinal *E. coli* infections in the studied community were caused by FZEC strains.

The public health implications of our findings are substantial. Since *E. coli* causes approximately 6 to 8 million UTIs in the U.S. annually [23], as many as 480,000 to 640,000 extraintestinal FZEC infections could occur in the U.S. each year. Our findings also implies that, collectively, FZEC strains could cause more UTIs annually than any non-*E. coli* uropathogenic species (e.g., *Klebsiella pneumoniae*) [24] or any of the major human-associated extraintestinal pathogenic *E. coli* lineages, including ST131-*H*30, ST95, and ST73.

Our study was conducted using isolates from Flagstaff Arizona – a small American city with high-quality infrastructure for potable water, sanitation, and hygiene, and no large-scale food-animal production. FZEC infections may be even more common in low- and middle-income countries, where the water, sanitation and hygiene infrastructure are less developed and where human populations live near food animals [25]. Likewise, the antimicrobial susceptibility of FZEC isolates may vary greatly with antimicrobial use patterns in different countries.

The present FZEC isolates differed significantly from both the meat associated *E. coli* populations and the non-FZEC clinical isolates with respect to antimicrobial susceptibility profiles, suggesting that they represented a distinct population. All the meat isolates were produced in the US; therefore, the *E. coli* populations that contaminate them should reflect domestic antimicrobial use practices. In contrast, the human population was a mix of residents of Flagstaff and surrounding communities as well as domestic and foreign visitors. Even the local human resident populations may have been exposed to diverse FZEC strains from foreign travel, which is a well-established risk factor for becoming colonized with antimicrobial-resistant Enterobacteriaceae [26].

Our novel MGE-based approach enabled us to quantify, with greater confidence than possible previously, the proportion of human extraintestinal *E. coli* infections caused by FZEC, and to identify high-risk lineages. Conventional molecular epidemiologic methods, including core genome phylogenetic analysis, have been critical for determining strain identity and for identifying point-source outbreaks [27–29]. By contrast, however, such methods are poorly suited for determining the host origins of sporadic infections, especially those involving highly diverse colonizing opportunistic pathogens such as *E. coli*.

Our analysis of *E. coli* populations from meat and humans indicates that certain sequence types are strongly associated with source. However, our detailed phylogenetic analysis suggests that even strains from strongly host-associated lineages can switch hosts. Therefore, relying on sequence type alone may lead to false assumptions about the origins of an isolate. We believe that the current best approach for predicting origins for a clinical *E. coli* isolate is to determine its lineage and source-associated MGE profile by whole genome sequencing and then applying our Bayesian model to generate a source probability.

An important benefit of our MGE-based approach is that, when used in conjunction with ST-specific core-genome phylogenies, it can differentiate recent host spillovers from historic host-switch events. Accurately identifying strains involved in recent animal-to-human spillovers is critical to developing targeted intervention strategies in food animal populations, e.g., vaccinating against high-risk lineages. By contrast,

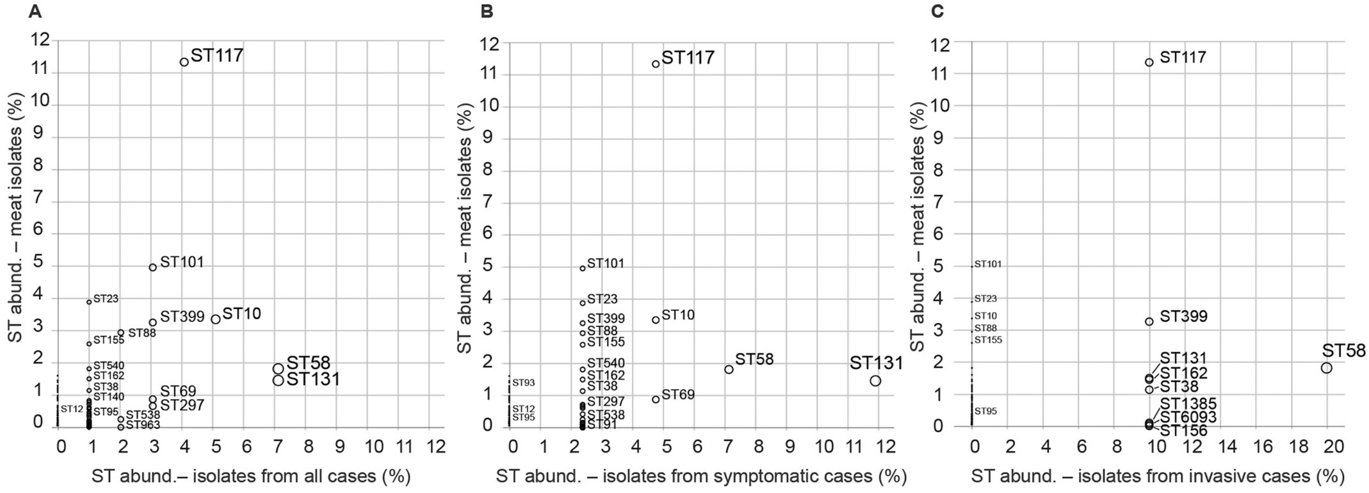

Fig. 4. Comparative abundance of *E. coli* sequence types (STs) among foodborne zoonotic *E. coli* (FZEC) isolates versus meat isolates. Panel A: all FZEC isolates; panel B: FZEC isolates from symptomatic cases; panel C: FZEC isolates from invasive cases. X-axes: ST abundance among FZEC isolates. Y-axes: ST abundance among meat isolates. Each ST is represented by an open circle, size-scaled for the proportion of FZEC isolates.

historic animal-to-human host-switch events are no longer actionable, because the strains involved have already become established within the human community.

Our study had some limitations. First, during model training we, of necessity, relied on sample type as a proxy for host origin even though our central hypothesis was that some level of incongruence exists between sample type and host origin. Despite this inescapable limitation, the model performed well according to the quality control metrics. Second, the study involved a single locale and time; therefore, the generalizability of the results is unknown. Third, since beef was not studied, we could not identify beef/cattle-associated MGEs, which may have caused us to underestimate FZEC cases. Finally, isolates from companion animal species, including cats and dogs, were not included in the study, which precluded us from evaluating their potential roles in the transmission of extraintestinal pathogenic *E. coli* strains.

Our study also had notable strengths. First, the sample set was a large, geographically localized collection of *E. coli* isolates, recovered contemporaneously from humans and retail meat products. Second, it used a rigorous comparative genomic approach, involving thousands of genomes, to identify MGEs that were differentially associated with *E. coli* isolates from meat and humans. Finally, its novel Bayesian latent class model combined the host-predictive values of these MGEs with an individual *E. coli* isolate's phylogenetic placement to objectively impute the isolate's source.

This study can serve as a foundation for future investigations, which could pursue multiple important objectives. One would be to expand the breadth of the model and improve its resolving power, thereby allowing assignment of isolates to specific vertebrate hosts rather than simply differentiating human from non-human. This conceivably could be done by enlarging and broadening both the isolate collection and the panel of host-associated accessory elements. Another possible objective would be to develop a model for estimating the elapsed time since a host transition took place, conceivably by determining the rate at which elements are lost or gained outside of selective hosts and incorporating this into the new model. Finally, a publicly available toolkit could conceivably be developed for investigators to use our approach to generate host-origin predictions for individual *E. coli* isolates based on genome sequences.

5. Conclusions

"My mama always said there's an awful lot you can tell about a person by their shoes. Where they're going. Where they've been." Zemeckis, R. *Forrest Gump*. Paramount Pictures. (1994).

Our study demonstrates the potential for using source-associated MGEs to infer the host origins of *E. coli* isolates. *E. coli* strains may shed or acquire host-adaptive MGEs in conjunction with host transitions. We liken this to people donning or doffing uniforms. Just as healthcare workers might switch their street clothes for scrubs and clogs for hospital work, an *E. coli* lineage might shed its avian-adaptive plasmids and acquire human-adaptive phage as it transitions from chickens to humans. And just as an observer would recognize people wearing scrubs and clogs as healthcare workers – even outside of a hospital context – one may be able to infer that a human clinical *E. coli* isolate carrying avian-adaptive MGEs was acquired recently from a poultry origin. Our novel approach can be applied in other settings to identify the highest-risk FZEC strains, determine their sources, and inform new one-health strategies to decrease the heavy public health burden imposed by extraintestinal *E. coli* infections.

Ethics approval and consent to participate

The Northern Arizona Healthcare IRB approved isolate collection and medical record review (protocol number: 573857–4) with a waiver of consent.

Consent for publication

NA

Availability of data and materials

All genome sequences have been uploaded to the NCBI SRA (accession numbers: PRJNA307689 and PRJNA407956).

Funding

This work was supported by the Wellcome Trust (201866); the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R01AI130066; the United States Army Medical Research and Materiel Command (W81XWH-11-1-0728); University of Michigan Precision Health Initiative; Northern Arizona University Cowden Endowment for Microbiology. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or other sponsors.

CRediT authorship contribution statement

Cindy M. Liu: Conceptualization, Formal analysis, Project administration, Writing – original draft. **Maliha Aziz:** Formal analysis, Visualization, Project administration, Writing – original draft. **Daniel E. Park:** Methodology, Visualization, Formal analysis, Writing – original draft. **Zhenke Wu:** Methodology, Resources, Formal analysis, Writing – review & editing. **Marc Stegger:** Formal analysis, Writing – original draft. **Mengbing Li:** Methodology, Formal analysis. **Yashan Wang:** Formal analysis, Visualization. **Kara Schmidlin:** Investigation, Supervision, Data curation. **Timothy J. Johnson:** Formal analysis, Writing – review & editing. **Benjamin J. Koch:** Conceptualization, Formal analysis, Writing – review & editing. **Bruce A. Hungate:** Conceptualization, Writing – review & editing. **Lora Nordstrom:** Data curation, Supervision. **Lori Gauld:** Data curation, Supervision. **Brett Weaver:** Data curation. **Diana Rolland:** Data curation. **Sally Statham:** Investigation. **Brantley Hall:** Formal analysis. **Sanjeev Sariya:** Formal analysis, Visualization. **Gregg S. Davis:** Formal analysis, Supervision, Funding acquisition. **Paul S. Keim:** Conceptualization, Formal analysis, Writing – original draft. **James R. Johnson:** Conceptualization, Formal analysis, Writing – original draft. **Lance B. Price:** Conceptualization, Formal analysis, Writing – original draft, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors report no conflicts of interest (see disclosure forms).

Data availability

Data will be made available on request.

Acknowledgements

The authors thank Joseph Horwinski and Amelia Dickens for their technical assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.onehlt.2023.100518.

References

[1] R.A. Shooter, E.M. Cooke, S.A. Rousseau, A.L. Breaden, Animal sources of common serotypes of *Escherichia coli* in the food of hospital patients. Possible significance in urinary-tract infections, Lancet 2 (1970) 226–228.

[2] E.M. Cooke, P.J. Kumar, R.A. Shooter, S.A. Rousseau, A.L. Foulkes, Hospital food as a possible source of *Escherichia coli* in patients, Lancet. 1 (1970) 436–437.

[3] R.A. Shooter, E.M. Cooke, M.C. Faiers, A.L. Breaden, S.M. O'Farrell, Isolation of *Escherichia coli*, *Pseudomonas aeruginosa*, and *Klebsiella* from food in hospitals, canteens, and schools, Lancet. 2 (1971) 390–392.

[4] I. Phillips, S. Eykyn, A. King, W.R. Gransden, B. Rowe, J.A. Frost, R.J. Gross, Epidemic multiresistant *Escherichia coli* infection in West Lambeth Health District, Lancet. 1 (1988) 1038–1041.

[5] S.J. Eykyn, I. Phillips, Community outbreak of multiresistant invasive *Escherichia coli* infection, Lancet. 2 (1986) 1454.

[6] A.R. Manges, J.R. Johnson, B. Foxman, T.T. O'Bryan, K.E. Fullerton, L.W. Riley, Widespread distribution of urinary tract infections caused by a multidrug-resistant *Escherichia coli* clonal group, N. Engl. J. Med. 345 (2001) 1007–1013.

[7] M. Ramchandani, A.R. Manges, C. DebRoy, S.P. Smith, J.R. Johnson, L.W. Riley, Possible animal origin of human-associated, multidrug-resistant, uropathogenic *Escherichia coli*, Clin. Infect. Dis. 40 (2005) 251–257.

[8] M. Bonten, J.R. Johnson, A.H.J. van den Biggelaar, L. Georgalis, J. Geurtsen, P.I. de Palacios, S. Gravenstein, T. Verstraeten, P. Hermans, J.T. Poolman, Epidemiology of *Escherichia coli* bacteremia: a systematic literature review, Clin. Infect. Dis. 72 (2021) 1211–1219.

[9] B. Foxman, The epidemiology of urinary tract infection, Nat. Rev. Urol. 7 (2010) 653–660.

[10] L.B. Price, B.A. Hungate, B.J. Koch, G.S. Davis, C.M. Liu, Colonizing opportunistic pathogens (COPs): the beasts in all of us, PLoS Pathog. 13 (2017), e1006369.

[11] T.J. Laird, D. Jordan, Z.Z. Lee, M. O'Dea, M. Stegger, A. Truswell, S. Sahibzada, R. Abraham, S. Abraham, Diversity detected in commensals at host and farm level reveals implications for national antimicrobial resistance surveillance programmes, J. Antimicrob. Chemother. 77 (2022) 400–408.

[12] T. Wirth, D. Falush, R. Lan, F. Colles, P. Mensa, L.H. Wieler, H. Karch, P.R. Reeves, M.C.J. Maiden, H. Ochman, M. Achtman, Sex and virulence in *Escherichia coli*: an evolutionary perspective, Mol. Microbiol. 60 (2006) 1136–1151.

[13] E. Denamur, O. Clermont, S. Bonacorsi, D. Gordon, The population genetics of pathogenic *Escherichia coli*, Nat. Rev. Microbiol. 19 (2021) 37–54.

[14] O. Tenaillon, D. Skurnik, B. Picard, E. Denamur, The population genetics of commensal *Escherichia coli*, Nat. Rev. Microbiol. 8 (2010) 207–217.

[15] K.A. Jolley, J.E. Bray, M.C.J. Maiden, Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications, Wellcome Open Res. 3 (2018) 124.

[16] S.K. Tiwari, B.C.L. van der Putten, T.M. Fuchs, T.N. Vinh, M. Bootsma, R. Oldenkamp, R. La Ragione, S. Matamoros, N.T. Hoa, C. Berens, J. Leng, J. Álvarez, M. Ferrandis-Vila, J.M. Ritchie, A. Fruth, S. Schwarz, L. Domínguez, M. Ugarte-Ruiz, A. Bethe, C. Huber, V. Johanns, I. Stamm, L.H. Wieler, C. Ewers, A. Fivian-Hughes, H. Schmidt, C. Menge, T. Semmler, C. Schultsz, Genome-wide association reveals host-specific genomic traits in *Escherichia coli*, bioRxiv (2022), https://doi.org/10.1101/2022.02.08.479532, 2022.02.08.479532.

[17] S.-C. Park, K. Lee, Y.O. Kim, S. Won, J. Chun, Large-scale genomics reveals the genetic characteristics of seven species and importance of phylogenetic distance for estimating pan-genome size, Front. Microbiol. 10 (2019) 834.

[18] C.M. Liu, M. Stegger, M. Aziz, T.J. Johnson, K. Waits, L. Nordstrom, L. Gauld, B. Weaver, D. Rolland, S. Statham, J. Horwinski, S. Sariya, G.S. Davis, E. Sokurenko, P. Keim, J.R. Johnson, L.B. Price, *Escherichia coli* ST131-*H*22 as a foodborne uropathogen, MBio. 9 (2018), https://doi.org/10.1128/mBio.00470-18.

[19] S.L. Jørgensen, M. Stegger, E. Kudirkiene, B. Lilje, L.L. Poulsen, T. Ronco, T. Pires Dos Santos, K. Kiil, M. Bisgaard, K. Pedersen, L.K. Nolan, L.B. Price, R.H. Olsen, P. S. Andersen, H. Christensen, Diversity and population overlap between avian and human *Escherichia coli* belonging to sequence type 95, mSphere 4 (2019), https://doi.org/10.1128/mSphere.00333-18.

[20] G.S. Davis, K. Waits, L. Nordstrom, B. Weaver, M. Aziz, L. Gauld, H. Grande, R. Bigler, J. Horwinski, S. Porter, M. Stegger, J.R. Johnson, C.M. Liu, L.B. Price, Intermingled *Klebsiella pneumoniae* populations between retail meats and human urinary tract infections, Clin. Infect. Dis. 61 (2015) 892–899, https://doi.org/10.1093/cid/civ428.

[21] CLSI, in: Performance standards for antimicrobial testing, CLSI Document M100-S24, 2014.

[22] M. Li, D.E. Park, M. Aziz, C.M. Liu, L.B. Price, Z. Wu, Integrating sample similarities into latent class analysis: a tree-structured shrinkage approach, Biometrics. (2021) 1–16, https://doi.org/10.1111/biom.13580.

[23] T.A. Russo, J.R. Johnson, Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem, Microbes Infect. 5 (2003) 449–456.

[24] A.L. Flores-Mireles, J.N. Walker, M. Caparon, S.J. Hultgren, Urinary tract infections: epidemiology, mechanisms of infection and treatment options, Nat. Rev. Microbiol. 13 (2015) 269–284.

[25] M.L. Nadimpalli, M. Stegger, R. Viau, V. Yith, A. de Lauzanne, N. Sem, L. Borand, B.-T. Huynh, S. Brisse, V. Passet, S. Overballe-Petersen, M. Aziz, M. Gouali, J. Jacobs, T. Phe, B.A. Hungate, V.O. Leshyk, A.J. Pickering, F. Gravey, C.M. Liu, T. J. Johnson, S. Le Hello, L.B. Price, Leakiness at the human-animal interface in Southeast Asia and implications for the spread of antibiotic resistance, bioRxiv (2021), https://doi.org/10.1101/2021.03.15.435134, 2021.03.15.435134.

[26] I. Frost, T.P. Van Boeckel, J. Pires, J. Craig, R. Laxminarayan, Global geographic trends in antimicrobial resistance: the role of international travel, J. Travel Med. 26 (2019), https://doi.org/10.1093/jtm/taz036.

[27] M.W. Allard, E. Strain, D. Melka, K. Bunning, S.M. Musser, E.W. Brown, R. Timme, Practical value of food pathogen traceability through building a whole-genome sequencing network and database, J. Clin. Microbiol. 54 (2016) 1975–1983.

[28] S. Quainoo, J.P.M. Coolen, S.A.F.T. van Hijum, M.A. Huynen, W.J.G. Melchers, W. van Schaik, H.F.L. Wertheim, Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis, Clin. Microbiol. Rev. 30 (2017) 1015–1063.

[29] S.D. Bentley, J. Parkhill, Genomic perspectives on the evolution and spread of bacterial pathogens, Proc. Biol. Sci. 282 (2015) 20150488.