Open Forum Infectious Diseases

MAJOR ARTICLE

Genomic Epidemiology of Major Extraintestinal Pathogenic *Escherichia coli* Lineages Causing Urinary Tract Infections in Young Women Across Canada

Chad D. Fibke,[1] Matthew A. Croxen,[2,3] Hyun Min Geum,[1,a] Magdalena Glass,[1,b] Eugenia Wong,[4,c] Brent P. Avery,[5] Danielle Daignault,[6] Michael R. Mulvey,[7,8] Richard J. Reid-Smith,[5,9] E. Jane Parmley,[5,9] Andrea Portt,[10] Patrick Boerlin,[11] and Amee R. Manges[1,12]

[1]School of Population and Public Health, The University of British Columbia, Vancouver, BC, Canada, [2]Department of Laboratory Medicine & Pathology, University of Alberta, Edmonton, AB, Canada, [3]Public Health Laboratory (ProvLab), Edmonton, AB, Canada, [4]Department of Microbiology and Immunology, The University of British Columbia, Vancouver, BC, Canada, [5]Center for Food-borne, Environmental and Zoonotic Infectious Disease, Public Health Agency of Canada, Guelph, ON, Canada, [6]National Microbiology Laboratory, Public Health Agency of Canada, Saint-Hyacinthe, QC, Canada, [7]Department of Medical Microbiology, University of Manitoba, Winnipeg, MB, Canada, [8]National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB, Canada, [9]Department of Population Medicine, Ontario Veterinary College, University of Guelph, Guelph, ON, Canada, [10]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada, [11]Department of Pathobiology, Ontario Veterinary College, University of Guelph, Guelph, ON, Canada, and [12]British Columbia Centre for Disease Control, Vancouver, BC, Canada

Background. A few extraintestinal pathogenic *Escherichia coli* (ExPEC) multilocus sequence types (STs) cause the majority of community-acquired urinary tract infections (UTIs). We examine the genomic epidemiology of major ExPEC lineages, specifically factors associated with intestinal acquisition.

Methods. A total of 385 women with UTI caused by *E. coli* across Canada were asked about their diet, travel, and other exposures. Genome sequencing was used to determine both ST and genomic similarity. Logistic regression was used to identify factors associated with the acquisition of and infection with major ExPEC STs relative to minor ExPEC STs.

Results. ST131, ST69, ST73, ST127, and ST95 were responsible for 54% of all UTIs. Seven UTI clusters were identified, but genomes from the ST95, ST127, and ST420 clusters exhibited as few as 3 single nucleotide variations across the entire genome, suggesting recent acquisition. Furthermore, we identified a cluster of UTIs caused by 6 genetically-related ST1193 isolates carrying mutations in *gyrA* and *parC*. The acquisition of and infection with ST69, ST95, ST127, and ST131 were all associated with increased travel. The consumption of high-risk foods such as raw meat or vegetables, undercooked eggs, and seafood was associated with acquisition of and infection with ST69, ST127, and ST131, respectively.

Conclusions. Reservoirs may aid in the dissemination of pandemic ExPEC lineages in the community. Identifying ExPEC reservoirs may help prevent future emergence and dissemination of high-risk lineages within the community setting.

Keywords. antimicrobial resistance; extraintestinal pathogenic *Escherichia coli*; food safety; molecular epidemiology; urinary tract infections.

Urinary tract infections (UTIs) are one of the most common bacterial infections among young women worldwide. It is estimated that a third of women will experience at least 1 UTI requiring antimicrobial treatment by the age of 24 [1]. Extraintestinal pathogenic *Escherichia coli* (ExPEC) are responsible for >80% of UTIs [2]. The intestinal acquisition of ExPEC has been associated with multiple exposures, including poultry contact or consumption, contact with companion animals, and direct transmission between sex partners and household contacts [3]. Once ExPEC colonize the intestinal tract, they can remain there for months to years without any ill effects [4]. Host behavior, medical interventions, physiologic abnormalities, and ExPEC virulence profiles facilitate transit from the gastrointestinal tract to extraintestinal body sites where they can cause infection [1]. These extraintestinal body sites include the meninges, bloodstream, surgical, abdominal cavity, prostate, respiratory, and most commonly the urinary tract [1, 2].

Multilocus sequence typing (MLST) is a standard approach for grouping evolutionarily related ExPEC into sequence types (STs) [5]. Despite the large diversity of ExPEC STs, recent studies show that ST12, ST69, ST73, ST95, ST127, and ST131 are predominant lineages causing extraintestinal infections worldwide [6–8]. Certain STs are also associated with specific antimicrobial resistance phenotypes. ST69 is associated with trimethoprim-sulfamethoxazole resistance [9], and ST131 is

Received 6 August 2019; editorial decision 19 September 2019; accepted 4 October 2019.
Present affiliations:
[a]Department of Medicine, The University of British Columbia, Vancouver, BC, Canada,
[b]Canada's Michael Smith Genome Sciences Centre, Vancouver, BC, Canada, and
[c]Gillings School of Global Public Health, The University of North Carolina, Chapel Hill, NC, USA
Correspondence: A. R. Manges, MPH, PhD, UBC, School of Population and Public Health, 2206 East Mall, Vancouver, BC V6T 1Z3 Canada (amee.manges@ubc.ca).

Open Forum Infectious Diseases®

DOI: 10.1093/ofid/ofz431

associated with extended-spectrum β-lactamases (ESBL) and fluoroquinolone resistance [10]. The pandemic ST131 lineage, specifically the H30R and H30Rx subgroups, has contributed to the global dissemination of cephalosporin and fluoroquinolone resistance [10]. In contrast, ST95, ST73, and ST127 are common causes of extraintestinal infections but are often fully susceptible.

We previously showed that prior antibiotic and UTI history, travel to Asia, consumption of chicken, shellfish, street foods, organic fruit, and contact with chickens, dogs, and pet treats are associated with UTIs caused by multidrug-resistant (MDR) ExPEC in women across Canada [11]. In this study, we extend our investigation and examine exposures associated with UTI caused by specific ExPEC lineages. We also investigate genomic relatedness between ExPEC isolates. Identifying the exposures and reservoirs of common, community-acquired ExPEC lineages will aid our understanding of the evolution, emergence, and dissemination of high-risk strains within the community setting.

METHODS

Study Design

This multisite epidemiologic study of women with community-acquired UTI has been described previously [11]. Research ethics boards at each university health clinic approved the study design (UBC REB H11-03439). Briefly, 399 women with suspected UTI caused by *E. coli* were enrolled from university health clinics at Saint Mary's University (SMU), McGill University (McGill), the University of Toronto (UofT), the University of Guelph (UofG), and The University of British Columbia (UBC) between 2012 and 2015. UTI was defined as the presence of at least 2 of the following symptoms: increased urinary frequency or urgency, dysuria, pyuria, or hematuria. Women who provided consent were asked to complete an online questionnaire requesting information on demographics, dietary habits, travel history, animal contact, history of UTI, health status, and other exposures over a 30-day period before their UTI. Only the initial *E. coli* isolate from women experiencing recurrent UTIs was included in the study.

E. coli Isolates and Antimicrobial Susceptibility

Isolate collection and antibiotic susceptibility testing have been described previously [11]. In brief, *E. coli* was transported using Uricults (Orion Diagnostica Uricult, Espoo, Finland). *E. coli* isolates were identified using indole testing and growth on both MacConkey and CHROMagar MH Orientation Chromogenic media (Becton-Dickinson, Paris, France). A single colony from the selective MacConkey agar was incubated overnight at 37°C in Lysogeny Broth and stored at −80°C in 15% glycerol. *E. coli* isolates were transferred to the National Microbiology Laboratory, Public Health Agency of Canada, Saint-Hyacinthe, Québec, for antimicrobial susceptibility testing by microdilution [11].

DNA Isolation and Sequencing

DNA was extracted from *E. coli* isolates using the PureLink Genomic DNA Mini Kit (Thermo Fisher) and used to construct sequencing libraries using the TruSeq DNA PCR Free Library Construction Protocol (Illumina). The first 161 *E. coli* isolates were sequenced at UBC Pharmaceutical Sciences Sequencing and Bioinformatics Consortium on the Illumina HiSeq 2500 platform. The second set of 241 *E. coli* isolates were sequenced at the British Columbia Genome Sciences Centre using the Illumina HiSeq X platform.

Genomic Sequence Analysis

Read quality was assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The Shovill 1.0 pipeline was used for both adaptor trimming and de novo genome assembly (https://github.com/tseemann/shovill). Adaptor trimming was performed with trimmomatic v0.36 [12] and genome assembly with SPAdes 3.12.0 [13]. The resulting contigs were kept if their lengths exceeded 500 bp and they had a minimum coverage of 3×. Draft genome quality was verified using taxonomic nearest-match identification with Mash v2.0 and RefSeq70 [14], contig metrics with QUAST v4.0 [15], and completeness with BUSCO v3.0 [16, 17]. Prokka v1.13.3 [18] and FimTyper v1.0 [19] were used for draft genome annotation and *fimH* typing, respectively.

Multilocus Sequence Typing Classification

Multilocus sequence typing (MLST) was performed in silico using the mlst v2.11 program (https://github.com/tseemann/mlst) using the Achtman scheme [20].

Identification of Acquired Antimicrobial Resistance Genes

The Comprehensive Antibiotic Resistance Database (CARD) Resistance Gene Identifier (RGI) was used to identify both acquired antimicrobial genes and mutations conferring resistance to an antimicrobial agent present in the genome assemblies [21]. Antimicrobial resistance genes were defined as present if they shared 95% identity, at 95% coverage, to a corresponding gene. Resistance gene classes were reported if at least 1 gene from a given class was present in 1 of the common ExPEC STs. Genes or gene classes present in all isolates are not shown.

Phylogenetic Analysis

The default settings in Roary v3.12.0 [22] were used to construct ST-specific core genome alignments. The resulting alignments were used for the creation of maximum likelihood phylogenetic trees using IQ-TREE v1.5.5 [23]. IQ-TREE's model finder [24] was employed to identify a consistent evolutionary model for all separate STs. The Generalized Time Reversible (GTR) model using FreeRate heterogeneity correction was chosen to configure each phylogeny. The minimal ancestor deviation method was used to root each phylogeny

[25]. One hundred bootstrap replicates were used for node support. The phylogenies were visualized with the ggtree package in R [26].

Single Nucleotide Variant Comparison

Snippy v4.3.0 (https://github.com/tseemann/snippy) was used to identify single nucleotide variants (SNVs) within each *E. coli* STs by read mapping [27, 28] and variant calling [29]. Mapping the reads of 1 isolate onto the entire draft genome of the corresponding isolate was performed in a pairwise manner.

Statistical Analysis

The outcome was defined as a UTI caused by 1 of the 5 major ExPEC STs, including ST69, ST73, ST95, ST127, and ST131. A single control group consisted of women with UTIs caused by other *E. coli* STs, including novel STs, singletons STs, and STs that appeared in <3% of women. Logistic regression was performed for each specific ST separately to examine exposures associated with major ExPEC ST acquisition and infection. Odd ratios, 95% confidence intervals (CIs), *P* values, and false discovery rate (FDR)–adjusted *P* values were visualized using

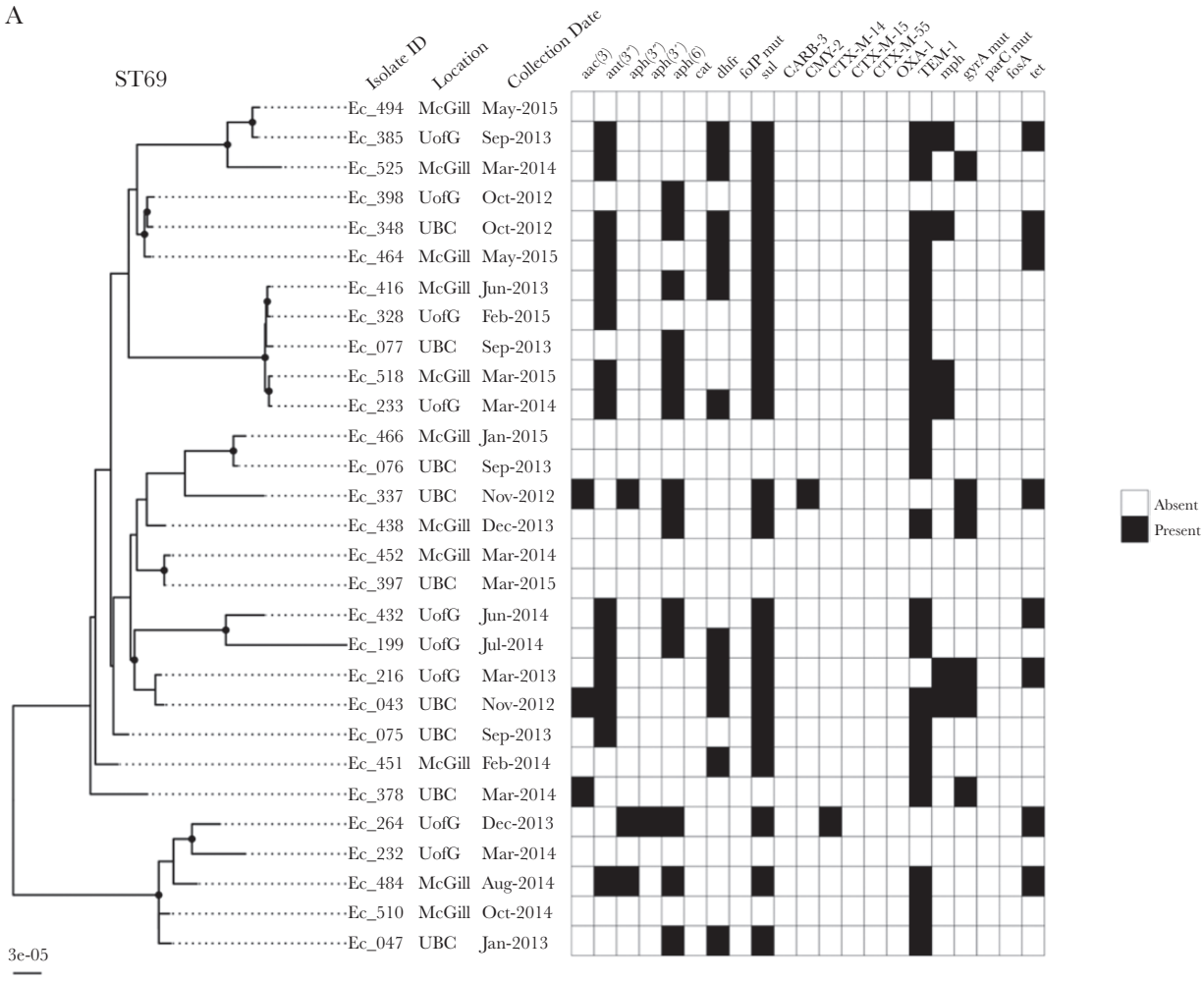

Figure 1. Core genome-based phylogeny of the 5 major ExPEC sequence types. A maximum likelihood core genome phylogeny was constructed for (A) sequence type (ST) 69, (B) ST73, (C) ST95, (D) ST127, and (E) ST131 using the GTR + Free Rate corrections evolutionary model with IQ-TREE (GTR + R2). Each tree was rooted using the minimal ancestral deviation method. Inner nodes supported with >90 of the 100 bootstrap replicates are represented as black points. The scale bar represents the expected number of nucleotide substitutions per site. Gray highlight denotes isolates that show genetic, temporal, and geographic relatedness. Location names are abbreviated as follows: Saint Mary's University (SMU), McGill University (McGill), the University of Toronto (UofT), University of Guelph (UofG), and The University of British Columbia (UBC). The ST131 phylogeny was further annotated with *fimH* type and the quinolone resistance phenotype of the isolate. The * for *fimH* complex denotes unknown *fimH* type. Acquired resistance genes were identified using resistance gene identifier (RGI) and the Comprehensive Antibiotic Resistance Database (CARD). Genes were grouped into gene families except for β-lactamase genes, the *fosA* gene and mutated dihydropteroate synthase (*folP*), gyrase subunit A (*gyrA*), and DNA topoisomerase 4 subunit A (*parC*) genes conferring resistance. Gene families were abbreviated as follows: aminoglycoside (3) acetyltransferase (*aac(3)*), aminoglycoside nucleotidyltransferase genes (*ant(3'')*), aminoglycoside phosphoryltransferase genes (*aph(3'); (aph(3''); aph(6)*), cephamycinase genes (CMY-2), β-lactamases active on cefotaxime (CTX-M-*), OXA β-lactamases (OXA-1), TEM β-lactamases (TEM-*), CARB β-lactamases (CARB-3), chloramphenicol acetyltransferase genes (*cat*), dihydrofolate reductase genes (*dhfr*), macrolide phosphotransferase genes (*mph*), tetracycline resistance genes (*tet*) including *tet(A); tet(B); tet(R),* and sulfonamide-resistant dihydropteroate synthase genes (*sul*).

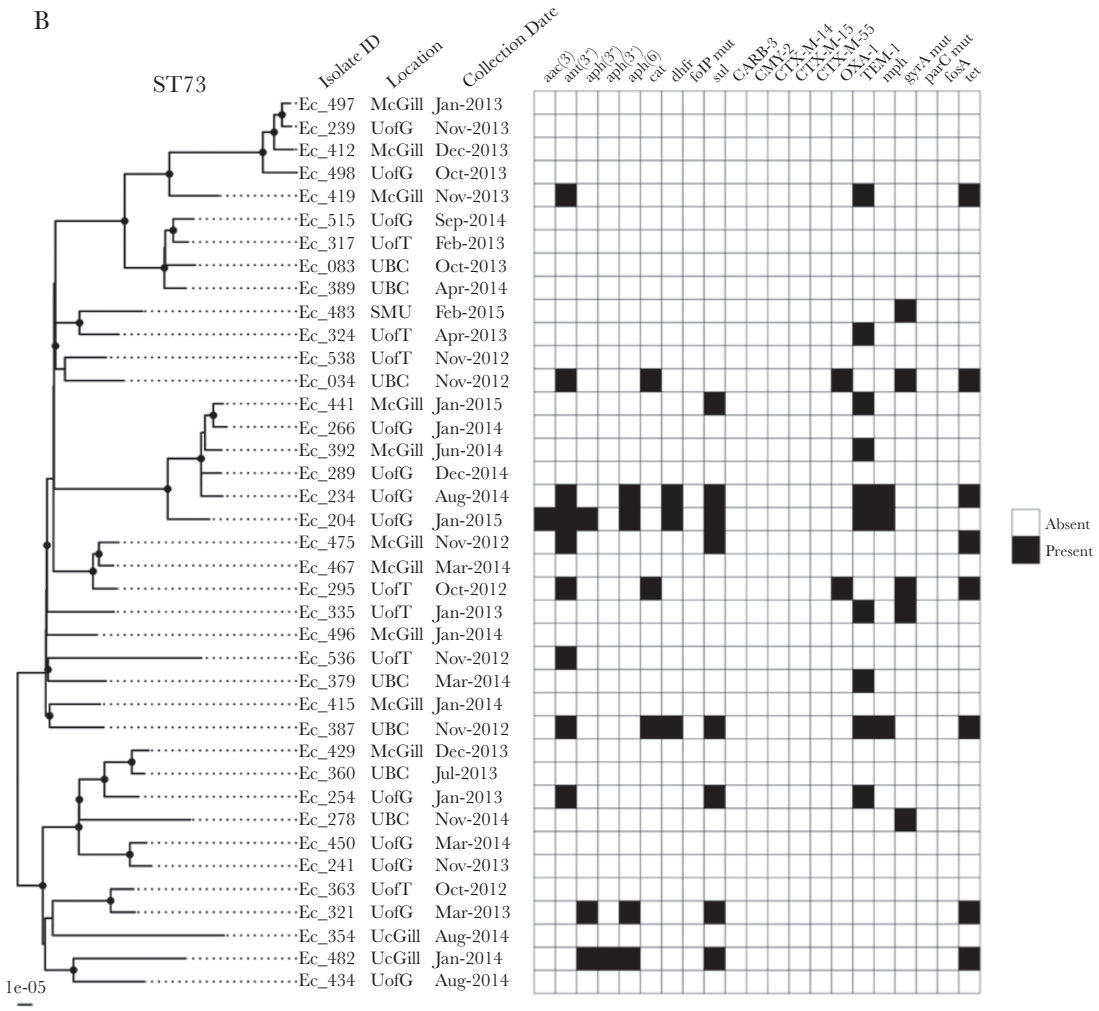

Figure 1. Continued

a Forest plot. Additional multiple logistic regression models were fit for ST69 and ST131, adjusting for single drug class resistance and MDR status, defined as resistance to ≥3 classes of antimicrobial agents. MDR adjustment would help determine whether the exposures were associated with genotype or resistance phenotype. All statistical analyses and visualization were performed with R, version 3.5.1.

RESULTS

A total of 399 women experiencing a UTI from October 2012 to May 2015 were included in this study. The majority of women were recruited from UofG (n = 135; 33%), UBC (n = 114; 28.8%), McGill (n = 113; 28.6%), and UofT (n = 32; 8.3%); 5 women were enrolled from SMU (1.3%).

E. coli Isolates and MLST Designation

Of the 399 sequencing data sets, 14 isolates were removed due to sequencing failure or poor genome metrics. *E. coli* genomes from 385 women with UTI were available for further analysis (Supplementary Data). Assembled genomes were classified into 87 different STs, whereas 12 isolates exhibited novel STs (Supplementary Data). The major STs included ST95 (n = 71), ST73 (n = 39), ST127 (n = 35), ST131 (n = 34), and ST69 (n = 29) and accounted for 54% of all UTIs. The control group included 177 women with a UTI caused by singleton STs (n = 134), novel STs (n = 12), and less common STs (n = 31).

Phylogenetic Analysis and Pairwise SNV Differences

Core genome-based phylogenies were constructed to determine the genomic relatedness within each major ST case group (Figure 1). Genomes were further interrogated to quantify the number of SNV differences. ST69 (core genes = 3739), ST73 (core genes = 3822), and ST127 (core genes = 3857) were heterogenic (Figure 1A, B, D). However, a pair of ST69 isolates (Ec_452 vs Ec_397) showed 33 SNV differences, and a pair of ST127 isolates showed 3 SNV differences (Figure 1A, D). Only the ST127 pair was geographically and temporally linked; the isolates were both collected at UBC a month apart (Figure 1D;

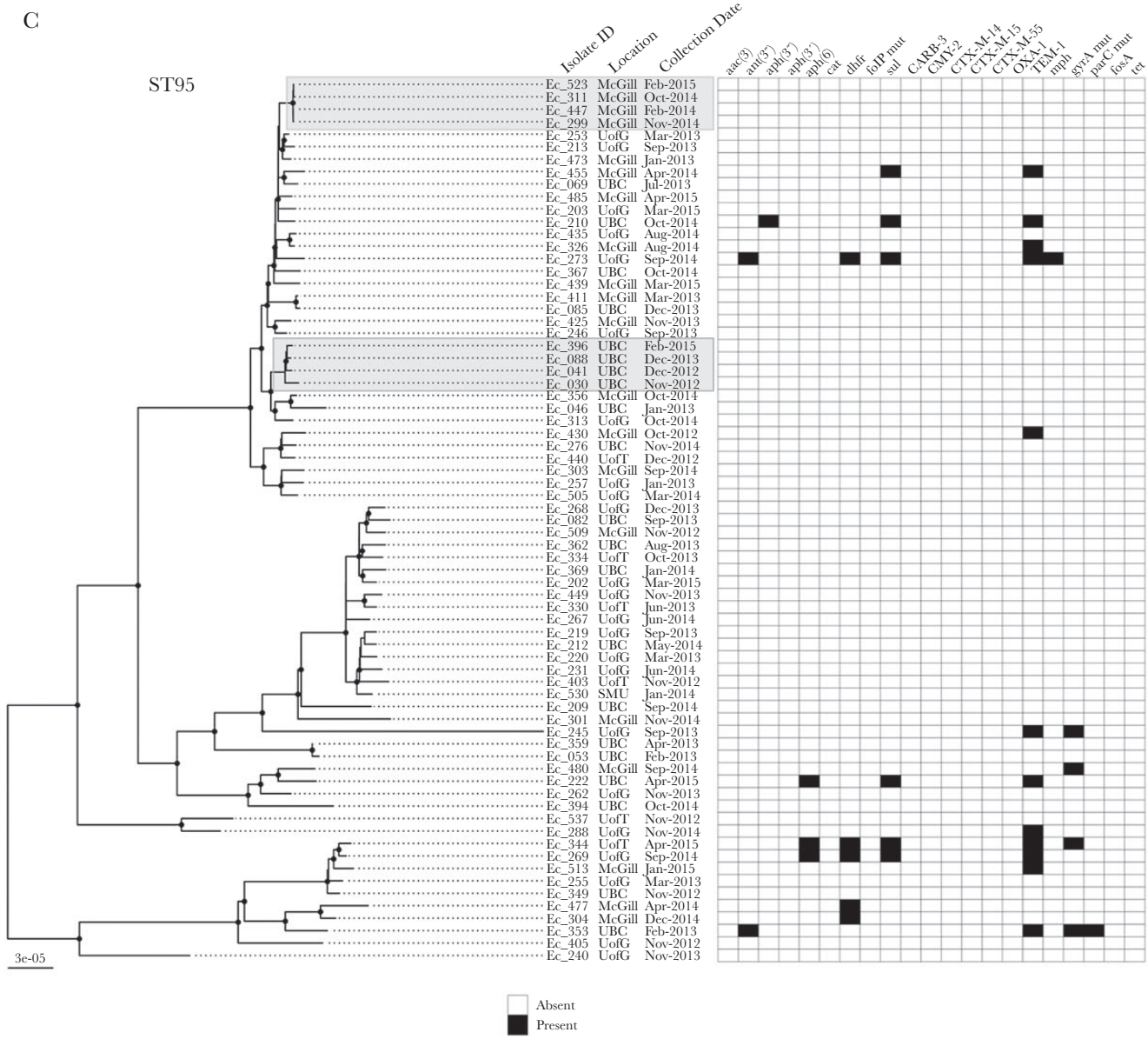

Figure 1. Continued

shaded section). The ST95 phylogeny (core genes = 3729) revealed a cluster from UBC exhibiting 37–73 pairwise SNV differences and a cluster from McGill exhibiting 3–11 pairwise SNV differences over the entire genomes (Figure 1C; shaded sections). *E. coli* from the McGill cluster were recovered within 1 year of each other. The ST131 phylogeny (core genes = 3616) contained 3 isolates from McGill that varied by 108–195 SNVs (Figure 1E; shaded section), which were recovered a year apart. Six genetically related ST131 H30R isolates were recovered from several geographical locations over a 2-year period (Figure 1E).

Multiple clusters of UTIs caused by genetically related *E. coli* isolates were detected within the control group, including ST420, ST998, and ST1193. Phylogenies for ST420, ST998, and ST1193 are provided in Supplementary Figure 1. ST420 contained 2 isolates that differed by 4 SNVs, and both isolates were recovered from women at UofT within a 2-week period (Supplementary Figure 1A). A single cluster of ST998 isolates (62 SNV differences) was also observed (Supplementary Figure 1B). ST1193 contained 6 isolates that differed by between 37 and 104 SNVs; however, these UTIs occurred at 3 separate sites (Supplementary Figure 1C). Five of the 6 closely related ST1193 isolates were MDR, and all isolates harbored mutations in *gyrA* and *parC* conferring resistance to quinolones (Supplementary Figure 1C). Isolate genomes within all clusters exhibited >20× average sequencing depth.

Antimicrobial Susceptibility

The frequency of antimicrobial resistance by class and outcome group is summarized in the Supplementary Data 3. The

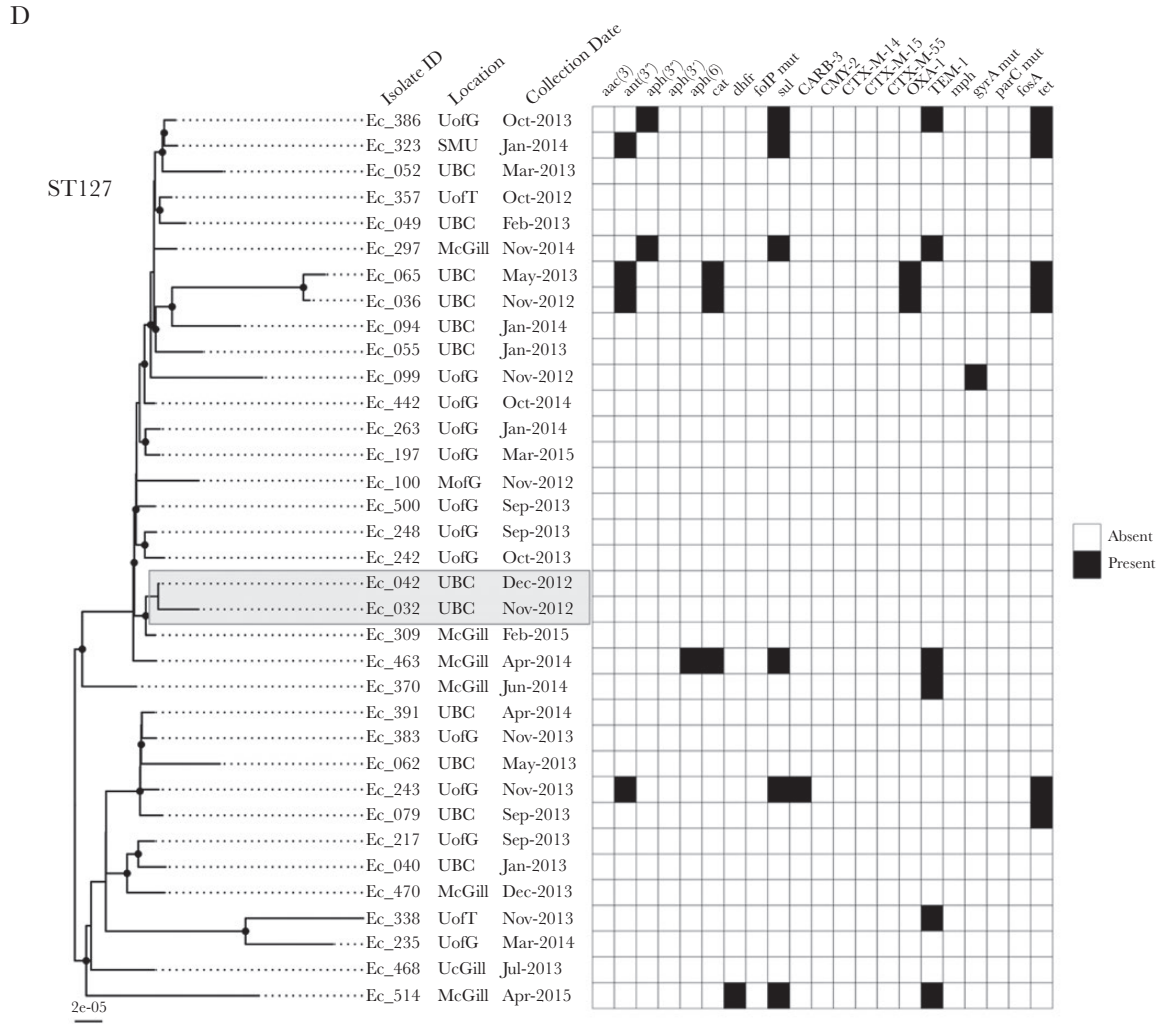

Figure 1. Continued

resistance phenotypes of ST127, ST95, and ST73 resembled the control group (Figure 2). Isolates belonging to ST69 and ST131 showed elevated odds of resistance to multiple antibiotic classes (Figure 2). The ST69 lineage was strongly associated with sulfonamide resistance (odds ratio [OR], 9.35; 95% confidence interval [CI], 3.91–22.33). The ST131 lineage was strongly associated with resistance to the macrolide (OR, 8.92; 95% CI, 3.34–23.84), quinolone (OR, 5.86; 95% CI, 2.59–13.62), and penicillin (OR, 4.41; 95% CI, 2.03–9.56) drug classes compared to the control group (Figure 2).

Antimicrobial Resistance Gene Profiles in Common ExPEC Genomes

Each phylogeny (Figure 1A–E) includes a heatmap of acquired antimicrobial resistance (AMR) genes identified in the genome sequences. AMR genes were infrequent in ST95, ST73, and ST127 linages (Figure 1B–D). Both ST131 and ST69 (Figure 1A, E) exhibited the most AMR genes out of all major lineages. TEM-1 was the most prevalent broad-spectrum β-lactamase in all lineages except for ST127, which contained OXA-1 (Figure 1A–E). CTX-M genes were present in ST69 and ST131 lineages (Figure 1A, E). ST131 was the only major ST with a detectable *fosA* gene.

Exposures Associated With Predominant Sequence Types

The relationships between specific exposures and infection with each predominant ST are presented in Figure 3. The Supplementary Data contains the complete epidemiologic analysis results. Women with UTIs due to ST127 reported more frequent consumption of undercooked eggs, daily consumption of dining hall food, and travel to Europe relative to the control group (Figure 3). Exposures negatively associated with UTI caused by ST127 were handling of seafood products and age (Figure 3).

Women with UTI caused by ST131 isolates reported more travel, higher consumption of uncooked shrimp, greater contact with fish, and daily consumption of peanut butter (Figure 3). Consumption of antibiotic-free meat was negatively associated with ST131 (Figure 3). After adjustment for multidrug resistance, consumption of fruit salad and tap water were linked

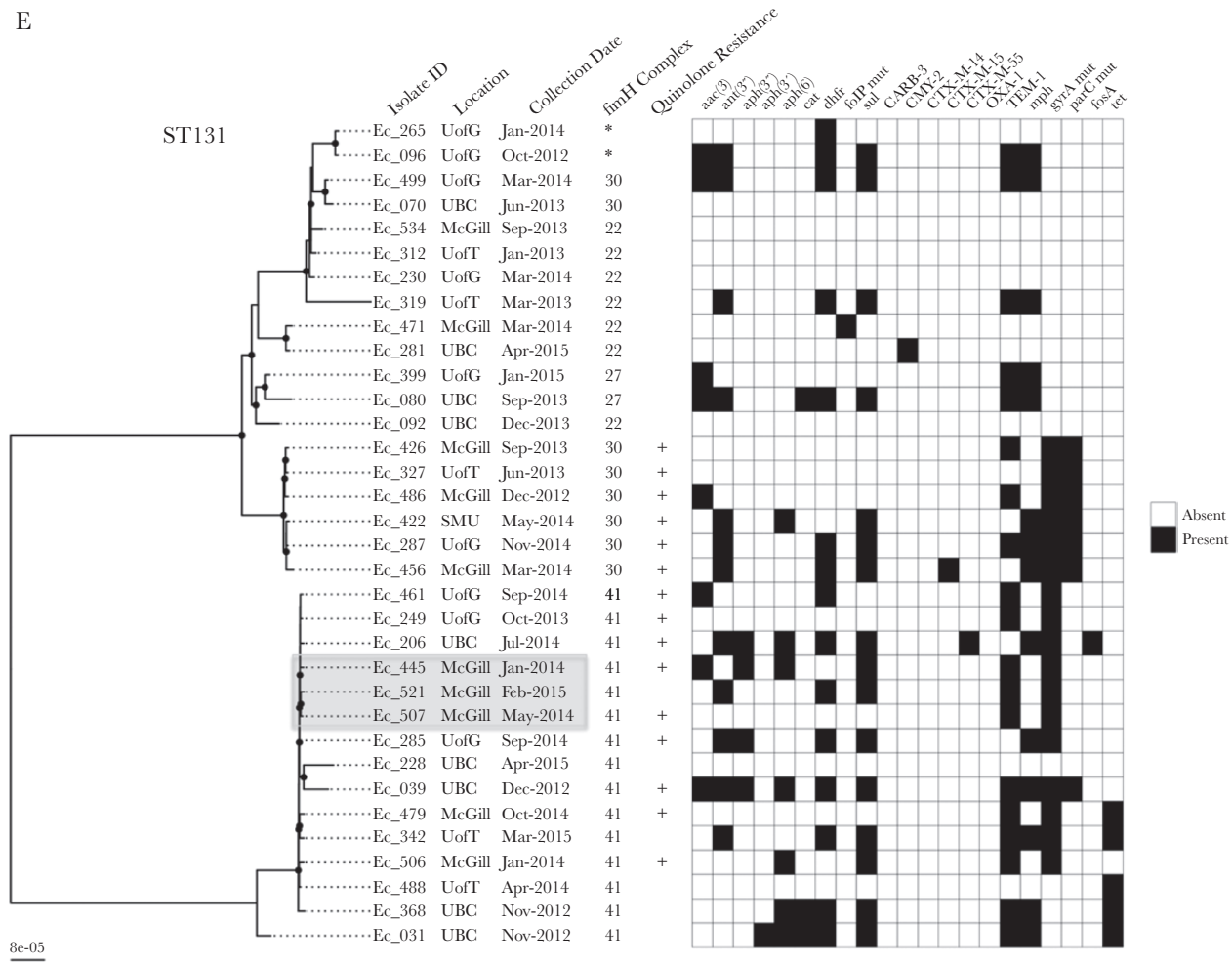

Figure 1. Continued

to ST131 UTI (Supplementary Data). Multidrug-resistant, sulfonamide- and macrolide-resistant ST131 isolates were associated with the consumption of shellfish more than 1–3 days per week (OR, 12.6; 95% CI, 1.9–25.2; OR, 27.0; 95% CI, 4.0–55.5; and OR, 5.7; 95% CI, 1.2–31.2, respectively).

Women with UTIs caused by ST69 reported more travel and greater consumption of raw meat, avocados, vegetable juice, organic dairy, and tap water than women in the control group (Figure 3). Ground beef preparation, consuming antibiotic-free meats, and drinking filtered tap water were negatively associated with ST69 UTI (Figure 3). Multidrug-resistant ST69 isolates were associated with the consumption of fruit smoothies >1–3 days per week (OR, 22.5; 95% CI, 3.07–271.1).

Women with UTI caused by ST73 reported more frequent consumption of street vendor–prepared food (Figure 3). Consumption of melons and eating at a student dining hall were negatively associated with a ST73 UTI (Figure 3).

Women with UTIs caused by ST95 reported more travel, greater peanut butter and almond consumption, and handling raw pet food compared with the control group (Figure 3). Consuming walnuts, using a Jacuzzi, eating other organic foods, and eating at restaurants, however, were negatively associated with UTI caused by ST95 (Figure 3).

DISCUSSION

In this study, we identified clusters of related ExPEC lineages and self-reported exposures associated with the acquisition of and infection with specific ExPEC lineages. ST95, ST127, ST131, ST73, and ST69 were the most common STs causing community-acquired UTIs in young women at 5 sites across Canada. This distribution of STs mirrors recent results from a study conducted at the University of California Berkeley between 1999 and 2017 [8]. These STs have been shown to be responsible for the greatest burden of extraintestinal infections worldwide [6, 7].

Core genome phylogenies of ST95, ST127, ST131, ST420, ST998, and STT1193 identified 7 genetically related isolate

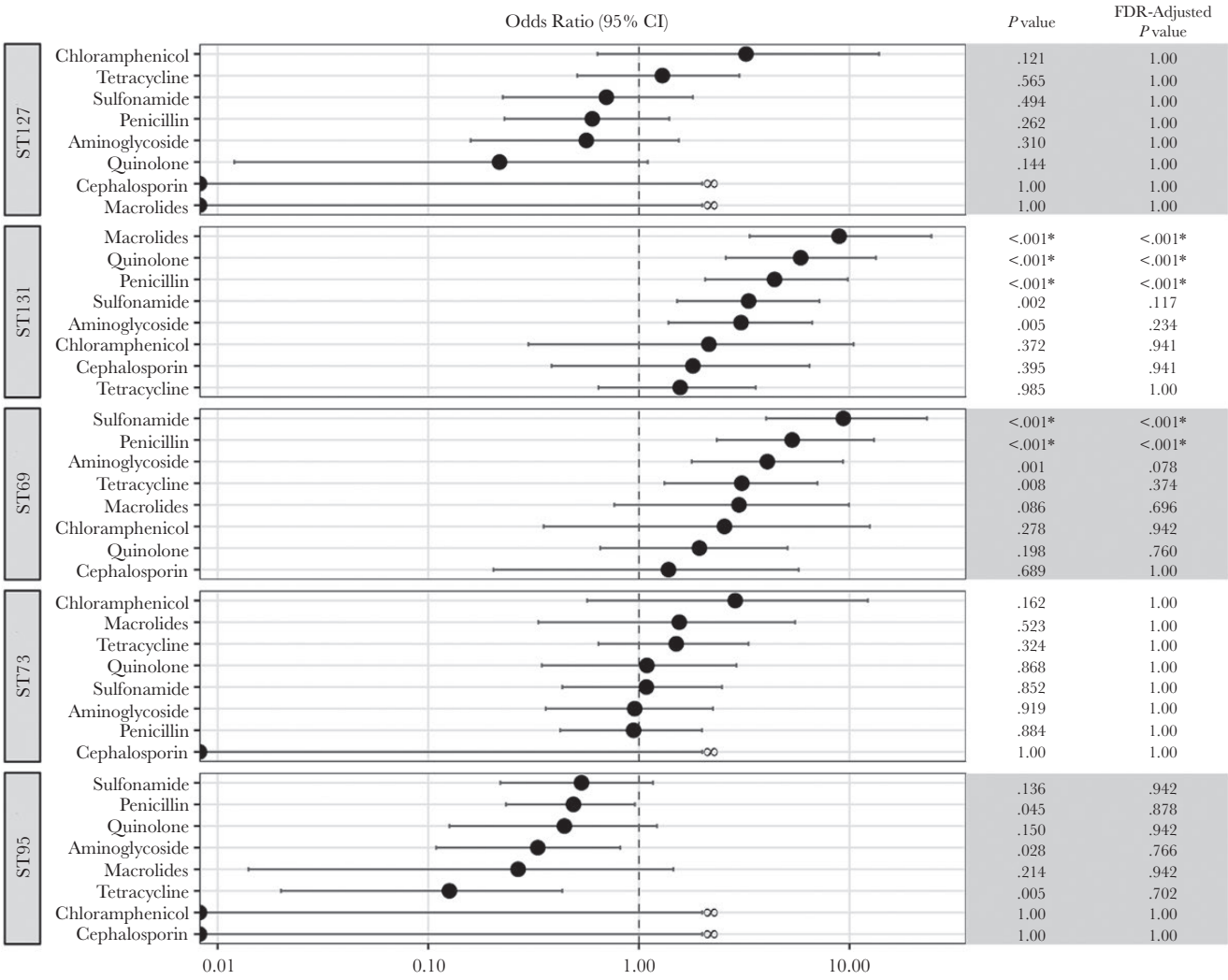

Figure 2. The relationship between major sequence types and class of antimicrobial resistance. Black circles represent the unadjusted odds ratio, and error bars represent 95% confidence intervals (CIs). The final column table shows the *P* value and false discovery rate (FDR)–adjusted *P* value for each estimate.

clusters. Furthermore, several ST95, ST127, and ST420 isolates exhibited nearly indistinguishable genomes. These genomic results resemble whole-genome sequencing-based analyses of enteric disease outbreak investigations [30]. ST95, ST420, ST998, ST1193, and ST131 clusters were observed but exhibited greater SNV differences. A cluster of UTIs caused by MDR ST1193 was observed between 2012 and 2014, which coincides with the emergence of fluoroquinolone-resistant ST1193 isolates in the United States [31].

ExPEC ST131 and ST69 were associated with greater resistance to multiple antimicrobial classes and carried the most CTX-M-related genes out of all predominant STs (Figures 1 and 2), which is consistent with prior investigations [32]. In contrast with other studies, we found that TEM-1 was more prevalent than CTX-M-15 [33]. This discrepancy may be explained by our inclusive sampling of all consecutive community-acquired UTIs. However, the increased prevalence of antimicrobial-susceptible ST73, ST95, and ST127 lineages suggests that factors other than resistance play a role in dissemination within the community.

Dissemination of these common ExPEC lineages may be related to consumption of established high-risk foods, travel, or other factors. ExPEC outbreaks have been identified across Europe and North America but these reports lack information on possible sources of the *E. coli* involved in the outbreaks [34]. Diarrheagenic *E. coli* outbreaks have been associated with the following high-risk foods: meat, fruit and vegetables, raw dairy products, raw eggs, and seafood or shellfish [35]. We show that all major ExPEC STs are associated with at least 1 of these high-risk foods (Figure 3). ST131 UTI is associated with self-reported consumption of shrimp and contact with fish. A recent study isolated several clinically relevant ExPEC STs from fish, including ST10, ST38, and ST131, which suggests that an aquatic reservoir may exist for ExPEC STs [36]. Women with ST131 or ST69 UTI reported less frequent consumption of

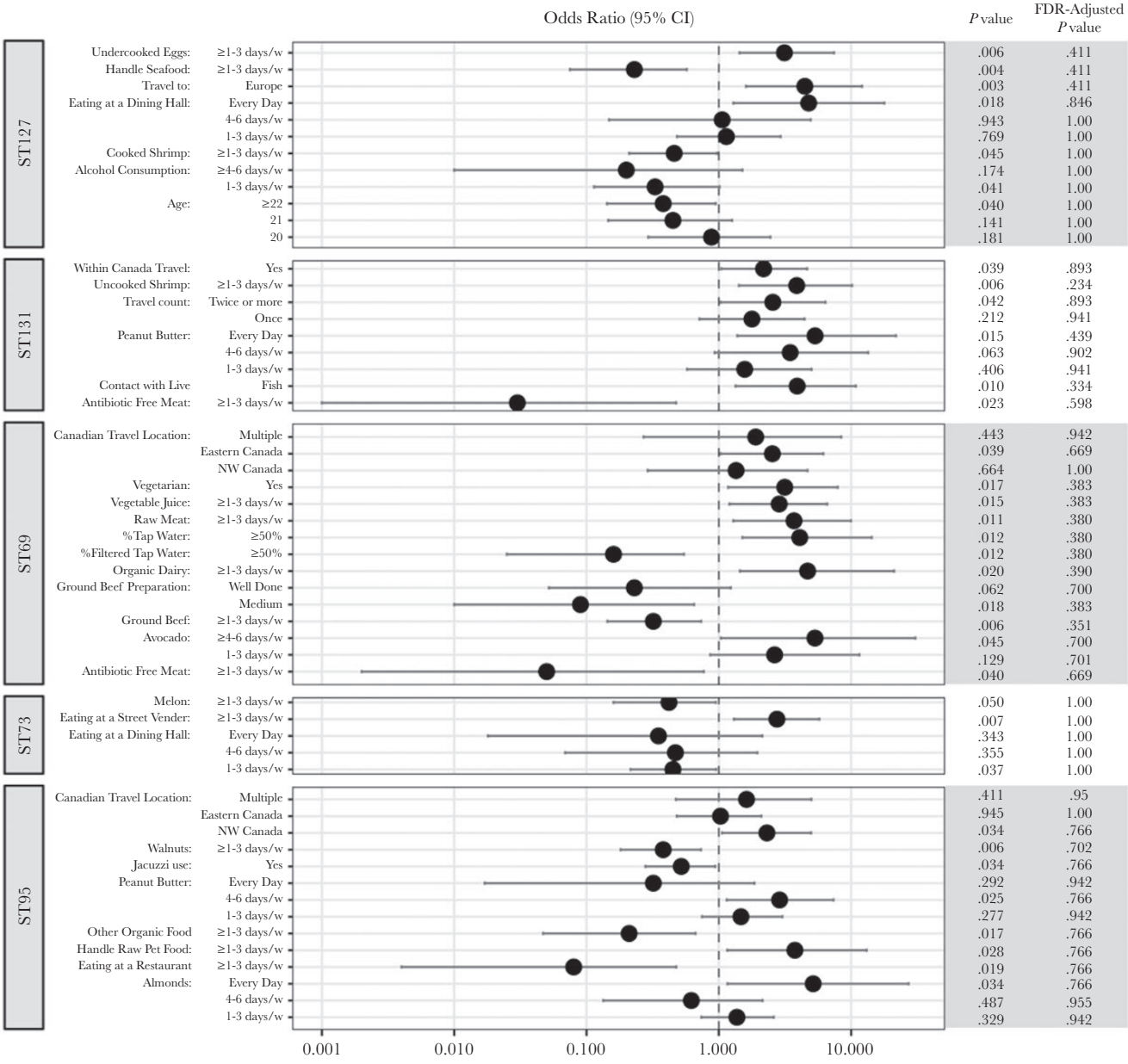

Figure 3. The relationship between risk factors and major ExPEC sequence types. Exposure variables are included if the corresponding unadjusted *P* value was <.05. Each reference group is defined as no exposure to the variable of interest. Black circles represent the estimated odds ratio, and error bars represent 95% confidence intervals (CIs). The final column table shows the *P* value and false discovery rate (FDR)–adjusted *P* value for each estimate. The reference category for age is <20 years, and the reference category for percent tap water and filtered tap water was <50%.

antibiotic-free meats. Both ST131 and ST69 ExPEC have been identified in food animal reservoirs [3]. Food animals exposed to antibiotics may select for ST131 and ST69 lineages, which could be transmitted to humans via retail meat. UTIs caused by ST69 were positively associated with consumption of raw meat and organic dairy, which have been associated with outbreaks caused by other *E. coli* pathotypes [35].

UTIs caused by ST73 and ST127 were associated with eating locations. These STs may be human host–specific STs and may be more likely to be transferred directly from person-to-person, as few studies have identified these STs in food or other animal reservoirs [3].

This study is the first to identify peanut butter consumption as a risk factor for UTI caused by both ST131 and ST95. A trace-back investigation was infeasible; therefore, a direct connection between these STs and exposure to peanut butter cannot be tested. However, a previous trace-back study confirmed hazelnuts to be the source of a multistate *E. coli* O157:H7 outbreak in the United States in 2011 [37]. It has been theorized that fecal contamination of fallen tree nuts arises from wild deer or cattle,

and in turn, the consumption of the contaminated tree nuts can lead to the acquisition of *E. coli* [37]. In addition to tree nuts, others have proposed the harvesting process as a likely source of *E. coli* contamination in the peanut's supply chain [38]. As the consumption of both almonds and peanut butter has been identified as risk factors for UTIs caused by ST95, it is plausible that some of the theorized routes of fecal contamination of tree nuts could apply to peanuts.

Travel variables were positively associated with UTIs caused by all major ExPEC lineages except for ST73. Travel to Europe was specifically associated with development of UTIs caused by ST127. ST127 is 1 of the dominant lineages in the United Kingdom [39]. It is possible that travel and/or travel-related disturbances to the intestinal microbiota make intestinal acquisition of new ExPEC strains more likely [40].

This study has several strengths and limitations. A small fraction of isolates had less optimal sequencing metrics, but this did not affect ST classification and cluster identification. We omitted dated phylogenies because of our short study period and concerns over phylogenetic dating modeling assumptions. Further limitations include the number of hypotheses tested, finite sample size, and limited statistical power. Future studies should also investigate person-to-person exposures that may be related to the acquisition of these major STs, especially for ST73. Poor recall of diet and other common exposures is frequent and will increase measurement error. Lastly, the control group included women with UTIs caused by less common, but clinically relevant STs, which may have influenced the identification of some relationships. Despite these limitations, consumption of high-risk foods and travel are consistent with intestinal acquisition of other *E. coli* pathotypes and previous studies investigating ExPEC reservoirs.

The study's strengths include the collection of all consecutive *E. coli* isolates causing community-acquired UTI from healthy women across Canada. This study is also one of the first to link epidemiological data with high-resolution genome sequencing data to address risk factors associated with developing UTIs caused by common ExPEC STs.

Investigation of reservoirs and transmission routes for pandemic ExPEC lineages is a public health priority, especially given the closely related genomes identified in this study. Furthermore, finding factors that contribute to the spread of successful lineages may help prevent the future emergence and dissemination of high-risk lineages within the community.

Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Acknowledgments

The authors would like to thank the study participants and clinic staff at Saint Mary's University, McGill University, and the University of Toronto, University of Guelph, and The University of British Columbia student health services.

Financial support. This work was supported by the Canadian Institutes of Health Research Canada Graduate Scholarship Master's Award (6556 to C.F.) and Canadian Institutes of Health Research, Institute of Infection and Immunity (MOP-114879 to A.R.M.).

Potential conflicts of interest. All authors: no reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

Data availability. Sequence data have been submitted to the NCBI's Sequence Read Archive under the BioProject PRJNA516477 (accession numbers: SRR8541263-SRR8541647).

References

1. Foxman B, Brown P. Epidemiology of urinary tract infections: transmission and risk factors, incidence, and costs. Infect Dis Clin North Am **2003**; 17:227–41.
2. Russo TA, Johnson JR. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. Microbes Infect **2003**; 5:449–56.
3. Johnson JR, Manges AR. Reservoirs of extraintestinal pathogenic *Escherichia coli*. Microbiol Spectr **2015**; 3:UTI-0006-2012.
4. Katouli M. Population structure of gut *Escherichia coli* and its role in development of extra-intestinal infections. Iran J Microbiol **2010**; 2:59–72.
5. Wirth T, Falush D, Lan R, et al. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol **2006**; 60:1136–51.
6. Riley LW. Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. Clin Microbiol Infect **2014**; 20:380–90.
7. Kallonen T, Brodrick HJ, Harris SR, et al. Systematic longitudinal survey of invasive *Escherichia coli* in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. Genome Res **2017**; 27:1437–49.
8. Yamaji R, Rubin J, Thys E, et al. Persistent pandemic lineages of uropathogenic *Escherichia coli* in a college community from 1999 to 2017. J Clin Microbiol **2018**; 56:e01834–17.
9. Johnson JR, Menard M, Johnston B, et al. Epidemic clonal groups of *Escherichia coli* as a cause of antimicrobial-resistant urinary tract infections in Canada, 2002 to 2004. Antimicrob Agents Chemother **2009**; 53:2733–9.
10. Nicolas-Chanoine MH, Bertrand X, Madec JY. *Escherichia coli* ST131, an intriguing clonal group. Clin Microbiol Rev **2014**; 27:543–74.
11. Ukah UV, Glass M, Avery B, et al. Risk factors for acquisition of multidrug-resistant *Escherichia coli* and development of community-acquired urinary tract infections. Epidemiol Infect **2018**; 146:46–57.
12. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **2014**; 30:2114–20.
13. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **2012**; 19:455–77.
14. Ondov BD, Treangen TJ, Melsted P, et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol **2016**; 17:132.
15. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics **2013**; 29:1072–5.
16. Simão FA, Waterhouse RM, Ioannidis P, et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **2015**; 31:3210–2.
17. Waterhouse RM, Seppey M, Simao FA, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol **2018**; 35:543–8.
18. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics **2014**; 30:2068–9.
19. Roer L, Tchesnokova V, Allesøe R, et al. Development of a web tool for *Escherichia coli* subtyping based on *fimH* alleles. J Clin Microbiol **2017**; 55:2538–43.
20. Jolley KA, Maiden MC. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics **2010**; 11:595.
21. Jia B, Raphenya AR, Alcock B, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res **2017**; 45:D566–73.
22. Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics **2015**; 31:3691–3.
23. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol **2015**; 32:268–74.
24. Kalyaanamoorthy S, Minh BQ, Wong TKF, et al. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods **2017**; 14:587–9.
25. Tria FDK, Landan G, Dagan T. Phylogenetic rooting using minimal ancestor deviation. Nat Ecol Evol **2017**; 1:0193.

26. Yu G, Smith DK, Zhu H, et al. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol **2017**; 8:28–36.
27. Li H, Handsaker B, Wysoker A, et al; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. Bioinformatics **2009**; 25:2078–9.
28. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv **2013**; Preprint at http://arxiv.org/abs/1303.3997.
29. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv **2012** Preprint at http://arxiv.org/abs/1207.3907
30. Pightling AW, Pettengill JB, Luo Y, et al. Interpreting whole-genome sequence analyses of foodborne bacteria for regulatory applications and outbreak investigations. Front Microbiol **2018**; 9:1482.
31. Johnson JR, Johnston BD, Porter SB, et al. Rapid emergence, subsidence, and molecular detection of *Escherichia coli* sequence type 1193- fimH 64 (ST1193- H 64), a new disseminated multidrug-resistant commensal and extraintestinal pathogen. J Clin Microbiol **2019**; 57:e01664-18.
32. Tourret J, Denamur E. Population phylogenomics of extraintestinal pathogenic *Escherichia coli*. Microbiol Spectr **2016**; 4. doi:10.1128/microbiolspec.UTI-0010-2012.
33. Smet A, Martel A, Persoons D, et al. Characterization of extended-spectrum β-lactamases produced by *Escherichia coli* isolated from hospitalized and nonhospitalized patients: emergence of CTX-M-15-producing strains causing urinary tract infections. Microb Drug Resist **2010**; 16:129–34.
34. George DB, Manges AR. A systematic review of outbreak and non-outbreak studies of extraintestinal pathogenic *Escherichia coli* causing community-acquired infections. Epidemiol Infect **2010**; 138:1679–90.
35. Yang SC, Lin CH, Aljuffali IA, Fang JY. Current pathogenic *Escherichia coli* foodborne outbreak cases and therapy development. Arch Microbiol **2017**; 199:811–25.
36. Bollache L, Bardet E, Depret G, et al. Dissemination of CTX-M-producing *Escherichia coli* in freshwater fishes from a French watershed (Burgundy). Front Microbiol **2019**; 10:3239.
37. Miller BD, Rigdon CE, Ball J, et al. Use of traceback methods to confirm the source of a multistate *Escherichia coli* O157:H7 outbreak due to in-shell hazelnuts. J Food Prot **2012**; 75:320–7.
38. Nascimento MS, Carminati JA, Silva ICRN, et al. *Salmonella*, *Escherichia coli* and *Enterobacteriaceae* in the peanut supply chain: from farm to table. Food Res Int **2018**; 105:930–5.
39. Ciesielczuk H, Jenkins C, Chattaway M, et al. Trends in ExPEC serogroups in the UK and their significance. Eur J Clin Microbiol Infect Dis **2016**; 35:1661–6.
40. Pitout JDD. Extraintestinal pathogenic *Escherichia coli*: a combination of virulence with antibiotic resistance. Front Microbiol **2012**; 3:9.