*Open Forum Infectious Diseases*

MAJOR ARTICLE

# Bacterial "Virulence" Traits and Host Demographics Predict *Escherichia coli* Colonization Behaviors Within Households

**Teresa C. Fox,[1,a] Connie Clabots,[2] Stephen B. Porter,[2] Tricia Bender,[2] Paul Thuras,[3,4] Aylin Colpan,[1,b] Jessica Boettcher,[1,c] and James R. Johnson[2,5]**

[1]Infectious Diseases, University of Minnesota, Minneapolis, Minnesota, USA, [2]Infectious Diseases, Veterans Affairs Medical Center, Minneapolis, Minnesota, USA, [3]Mental Health Patient Service Line, Veterans Affairs Medical Center, Minneapolis, Minnesota, USA, [4]Department of Psychiatry, University of Minnesota Minneapolis, Minnesota, USA, and [5]Infectious Diseases, University of Minnesota Minneapolis, Minnesota, USA

***Background.*** Although intestinal colonization precedes most extraintestinal *Escherichia coli* infections, colonization-promoting factors are incompletely understood. We compared within-household *E. coli* colonization patterns with host and bacterial traits.

***Methods.*** Twenty-two veterans with a clinical *E. coli* isolate and their 46 human and animal household members underwent longitudinal fecal sampling. Distinct *E. coli* strains were characterized for phylogenetic background, virulence genes, antibiotic resistance, and colonization behaviors. Host and bacterial traits were assessed statistically as predictors of colonization behaviors.

***Results.*** Among the 139 unique-by-household fecal *E. coli* strains, univariable predictors of colonization behavior included (i) host demographics, (ii) matching the index clinical isolate, and (iii) bacterial characteristics (2 phylogroups, 5 clonal lineages, 18 virulence genes, and molecular extraintestinal pathogenic *E. coli* status). Multivariable predictors of colonization behavior included veteran host, spouse host, matching the index clinical isolate, phylogroup F, ST73, *hlyD* (alpha hemolysin), *hlyF* (variant hemolysin), H7 *fliC* (flagellar variant), *vat* (vacuolating toxin), and *iha* (adhesin-siderophore).

***Conclusions.*** Host demographics, multiple bacterial "virulence" traits, and matching the index clinical isolate predicted *E. coli* fecal colonization behaviors. Thus, certain bacterial characteristics may promote both colonization and pathogenicity. Future interventions directed toward such traits might prevent *E. coli* infections both directly and by disrupting antecedent colonization.

**Keywords.** *Escherichia coli*; intestinal colonization; ST131-*H*30; strain sharing; virulence factors.

Extraintestinal *Escherichia coli* infections are an ever-growing threat that is compounded by emerging antimicrobial resistance. How antimicrobial-resistant *E. coli* strains emerge and disseminate is unclear. Because the fecal microbiota is a reservoir for such organisms, sharing of fecal *E. coli* strains among individuals likely contributes to dissemination [1–3].

Specific molecular traits—including certain phylogenetic groups, clonal lineages, and virulence genes—are over-represented among the extraintestinal pathogenic *E. coli* (ExPEC) strains that cause most extraintestinal *E. coli* infections. For example, sequence type (ST) 131 (from phylogenetic group B2), and specifically its ST131-*H*30 subclone, is the leading cause of fluoroquinolone-resistant and extended-spectrum beta-lactamase-producing *E. coli* infections worldwide [4, 5]. ExPEC strains commonly exhibit multiple virulence-associated accessory traits such as adhesins, siderophore systems, toxins, protectins, and motility mechanisms, some of which—although regarded as contributing to pathogenesis—may also promote colonization of human and animal hosts [6–21].

Despite the gut being the main reservoir for ExPEC, factors that affect colonization by and sharing of *E. coli* strains are incompletely defined. Accordingly, we evaluated *E. coli* colonization patterns within veterans' households and assessed host demographic subsets and bacterial traits as predictors of bacterial colonization behaviors.

## METHODS

### Study Population

This prospective observational cohort study involved veterans at the Minneapolis Veteran Affairs Medical Center (MVAMC) and their household members. Source veterans for sequential clinical *E. coli* isolates were sent a letter soliciting study participation, including any household members; veteran-only households were excluded. Eleven households, each with a fluoroquinolone-resistant or fluoroquinolone-susceptible index clinical isolate (hereafter, "fluoroquinolone-resistant households" and "fluoroquinolone-susceptible households"), were recruited, giving 22 total households. Of the 22 index isolates, 20 were from urine and 2 were from swabs; 15 represented

Received 14 April 2020; editorial decision 13 October 2020; accepted 18 October 2020.

[a]**Present affiliations:** Infectious Diseases, HealthPartners, St. Paul, Minnesota, USA, [b]Infectious Diseases, Lehigh Health Network, Allentown, Pennsylvania, USA, and [c]Infectious Diseases, Hennepin Health, Minneapolis, Minnesota, USA

Correspondence: Teresa C. Fox, MD, HealthPartners Specialty Center, 401 Phalen, Minneapolis, MN 55455 (foxx0243@umn.edu).

**Open Forum Infectious Diseases®2020**



DOI: 10.1093/ofid/ofaa495

infection, 4 represented colonization (urine, skin), and 3 were indeterminate (urine, wound). The 22 index subjects included 3 females and 19 males. Eighteen had notable comorbid health conditions, 13 reported antibiotic use in the past 6 months, and 6 reported antibiotic use in the past month (Supplementary Table 1).

#### Patient Consent Statement

The MVAMC Institutional Review Board approved the study before enrollment of participants. Written consent was obtained from all study participants.

#### Data Collection

The 68 total study participants (22 veterans, 27 human household members, 19 pets) underwent serial fecal sampling 2–6 times each between October 2012 and May 2014. Fecal samples were cultured on gram-negative selective agar, with and without ciprofloxacin (4 mg/L), to recover total and ciprofloxacin-resistant gram-negative bacilli. Morphologically consistent, indole-positive, citrate-negative colonies were regarded as *E. coli*. Up to 10 *E. coli* colonies per plate (as available) underwent random amplified polymorphic DNA (RAPD) analysis using primer 1254 according to an established procedure adapted from Berg et al. and optimized by the authors [22, 23]. In validation studies, when applied to replicate colonies from a given fecal sample, this approach yielded 85%–90% concordance with *Xbal* pulsed-field gel electrophoresis (PFGE) (unpublished, J.R.J.). One colony per unique RAPD profile per plate (as judged by visual inspection) and all index clinical isolates underwent standardized *Xbal* PFGE analysis [17, 24]. Pulsotypes were assigned based on 94% similarity to reference profiles within a large private PFGE library (2382 pulsotypes, 7092 *E. coli* isolates) [24]. This approach to resolving unique within-household shared and persisting colonizing strains has been validated by whole-genome sequencing [18].

#### Molecular Traits and Antibiotic Resistance

Each unique fecal strain (ie, pulsotype) per subject, plus the index isolates, underwent determination of major phylogenetic group (A, B1, B2, C, D, F), selected clonal lineages and subsets thereof (n = 15), and putative virulence genes (n = 47) with established multiplex polymerase chain reaction (PCR) assays [25, 26]. ExPEC status was defined operationally as presence of ≥2 of *papAH* and/or *papC* (P fimbriae), *sfa/focDE* (S and F1C fimbriae), *afa/draBC* (Dr-binding adhesins), *iutA* (aerobactin receptor), and *kpsM*II (group 2 capsules) [27]. The index isolates underwent molecular O typing using an established multiplex PCR assay that detects the 10 most common O types among invasive *E. coli* isolates [28]. Susceptibility to 24 common antibiotics was assessed by disk diffusion using Clinical and Laboratory Standards Institute–specified procedures and interpretive criteria. For strains with discrepant results among multiple representatives from a given household, the consensus result was used.

#### Colonization Outcome Definitions

For each unique *E. coli* strain per household, 4 colonization outcomes were determined. (i) The fecal predominance index (hereafter: fecal predominance) was the mean by-sample proportion of assessed *E. coli* colonies for which the strain accounted among those fecal samples that yielded the strain. (ii) The persistence index (hereafter: persistence) was the by-household mean proportion of a subject's fecal samples that yielded the strain, among samples collected after the strain's first detection in that subject. (iii) The sharing index (hereafter: sharing) was the proportion of potential sharing pairs within a household in which the strain was detected (synchronously or asynchronously) in both subjects. (iv) Household prevalence was the proportion of all of the household's fecal samples that yielded the strain.

#### Statistical Analysis

Variables associated with ≥5 fecal strains were analyzed statistically. For univariable analysis, comparisons involving dichotomous variables were assessed using the (n-1) chi-square test [29] or binomial logistic regression; those involving continuous variables were assessed using the Mann-Whitney test or simple regression. Due to their strongly bimodal distribution (not shown), persistence and sharing were analyzed as dichotomous variables.

Multivariable analysis was used to assess jointly host demographic subset, clinical isolate–matching status, and molecular traits as predictors of the studied colonization outcomes. For each colonization outcome, statistically significant univariable predictor variables were assessed for correlations, and clusters were identified of highly correlated variables (correlation coefficient ≥ 0.7). Within each such cluster, the variable most highly correlated with the particular outcome was selected for multivariable analysis. Backward stepwise variable entry was used primarily; forced and forward stepwise variable entry was used for sensitivity analysis.

The final multivariable analyses did not include antibiotic resistance phenotypes, given the study's main focus on the relationship between intrinsic bacterial characteristics (virulence genes, phylogenetic and clonal background) and colonization behaviors. In an exploratory multivariable analysis that included antibiotic resistance, streptomycin was the only resistance marker that significantly predicted any colonization outcome (household sharing: not shown).

Statistical analyses were performed using SPSS, version 19 (IBM Analytics). Throughout, the significance criterion was $P < .05$.

## RESULTS

### Index Clinical Isolates: Molecular Characteristics and Antibiotic Resistance

Of the 22 index clinical isolates, 18 were phylogroup B2; the rest were 1 each of phylogroups A, C, D, and F (Supplementary Table 2). ST and O-type could be determined for 18 of 21 and 16 of 21 index isolates, respectively. By ST (and O-type), in order of decreasing frequency, 6 isolates were ST131-*H*30R1 (O25b), 3 were ST73 (2 O25a, 1 O-undetermined), 2 were ST12 (O4), 2 were ST 95 (1 O2, 1 O1), and 1 each were ST1193 (O75), ST127 (O6), ST131-*H*30Rx (O25b), ST141 (O-undetermined), ST648 (O1), ST69 (O17). All but 1 index strain exhibited at least 1 virulence gene, and all but 2 showed phenotypic resistance to at least 1 of the 24 antibiotics (Supplementary Table 2).

### Household Category Comparisons: Fecal Sampling

The 22 study households comprised 68 total participants, including 49 humans (22 veterans, 27 others) and 19 pets. The 11 fluoroquinolone-resistant and 11 fluoroquinolone-susceptible households, so labeled based on the index veteran's clinical isolate, comprised, respectively, 31 and 37 total subjects (Table 1). Households contributed a mean (range) of 3.1 (2–10) subjects each. Overall, subjects provided a mean (range) of 3.4 (2–6) fecal samples each, and households provided a mean (range) of 9 (4–30) total samples each, for 208 total samples (100 for fluoroquinolone-resistant households; 108 for fluoroquinolone-susceptible households). Fecal samples were submitted over intervals ranging, by subject, from 5 to 62 weeks (median, 31 weeks).

Fecal samples yielded 139 unique-by-household *E. coli* strains. Of these, 16 (12%) matched the household's index clinical isolate (hereafter, "clinical isolate–matching strains"), whereas 123 (88%) did not (hereafter, "fecal-only strains").

### Household Category Comparisons

Microbiological results differed between the fluoroquinolone-resistant and fluoroquinolone-susceptible households (Table 2).

**Table 1. Household Composition in Relation to the Fluoroquinolone Phenotype of the Veteran's Index Clinical *Escherichia coli* Isolate**

| | Fluoroquinolone Phenotype of Index Clinical Isolate (No. per Category) | |
|---|---|---|
| Category | Resistant[a] | Susceptible[b] |
| Households | 11 | 11 |
| Total subjects | 31 | 37 |
| Veterans | 11 | 11 |
| Spouses | 9 | 10 |
| Nonspouse adults | 3 | 1 |
| Children | 0 | 4 |
| Pets | 8 | 11 |

[a]Households in which the veteran's index clinical *E. coli* isolate was fluoroquinolone-resistant.
[b]Households in which the veteran's index clinical *E. coli* isolate was fluoroquinolone-resistant.

ST131 accounted for the index clinical isolate for 7/11 (64%) fluoroquinolone-resistant households but no fluoroquinolone-susceptible household ($P < .001$). Collectively, fluoroquinolone-resistant households yielded numerically fewer unique-by-household fecal *E. coli* strains (n = 53) than did fluoroquinolone-susceptible households (n = 86; $P = .17$).

Colonization with a clinical isolate–matching strain was detected in proportionally more subjects from fluoroquinolone-resistant than fluoroquinolone-susceptible households, both among veterans (10/11 [91%] vs 5/11 [45%]; $P = .03$) and overall (17/31 [55%] from 10 households vs 9/37 [24%] from 6 households; $P = .01$) (Table 2). In veteran fecal samples, ST131 accounted for 6/10 (60%) of the clinical isolate–matching strains from fluoroquinolone-resistant households, but none from fluoroquinolone-susceptible households ($P = .04$).

As compared with fluoroquinolone-susceptible households, fluoroquinolone-resistant households yielded numerically fewer fecal-only strains (43 vs 80; $P = .14$) but, among those strains, a larger fluoroquinolone-resistant fraction (13/43 [30%] from 5 households vs 1/80 [1%] from a single household; $P < .001$) (Table 2). The fluoroquinolone-resistant households also yielded a larger fraction of ST131 fecal-only strains (7/43 [16%] from 5 households vs 1/80 [1%]; $P < .001$).

### Colonization Outcomes: Host Demographic Subset

Of the 139 total (unique-by-household) fecal strains, 63 occurred in veterans, 50 in spouses, 14 in children, and 43 in pets (some strains had multiple hosts). Host subset corresponded significantly with persistence, sharing, and/or household prevalence, but not fecal predominance (Table 3). Specifically, spouse source predicted persistence ($P = .01$), sharing ($P = .02$), and household prevalence ($P < .001$); veteran source predicted sharing ($P = .02$) and household prevalence ($P < .001$); and child source negatively predicted household prevalence ($P = .001$).

### Colonization Outcomes: Clinical Isolate Matching

Of the 16 total (unique-by-household) clinical isolate–matching fecal strains, 15 were identified in veterans, 7 in spouses, and 4 in pets. As compared with the 123 fecal-only strains, the 16 clinical isolate–matching fecal strains exhibited significantly greater persistence ($P < .001$), sharing ($P = .05$), and household prevalence ($P < .001$), but similar fecal predominance (Table 4).

### Colonization Outcomes: Molecular Traits

Three categories of molecular traits (phylogroups, clonal lineages, and virulence genes) were assessed as predictors of fecal colonization outcomes. By phylogroup, the fecal *E. coli* strains were distributed as follows (% of 139): B2, 48%; A, 21%; B1, 17%; D, 9%; F, 5%; and C, 1%. Phylogroups B2 and D ranked numerically highest for fecal predominance, persistence, and household prevalence, and for sharing were surpassed only by

**Table 2. Household *Escherichia coli* Fecal Strain Isolation in Relation to the Fluoroquinolone Phenotype of the Index Veteran's Clinical *E. coli* Isolate**

| Category | Fluoroquinolone Phenotype of Index Clinical Isolate (No. or Proportion) | | *P*[c] |
|---|---|---|---|
| | Resistant Households[a] | Susceptible Households[b] | |
| Total No. fecal strains[d] | 53 | 86 | |
| Total No. fecal-only strains[e] | 43 | 80 | |
| Fluoroquinolone-resistant fecal-only strains | 13/43 | 1/80 | <.001 |
| ST131 fecal-only strains | 7/43 | 1/80 | <.001 |
| Households with clinical isolate–matching strain in feces | 10/11 | 6/11 | |
| Subjects with clinical isolate–matching strain in feces | 17/31 | 9/37 | .01 |
| Veterans with clinical isolate–matching strain in feces | 10/11 | 5/11 | .03 |
| Spouses with clinical isolate–matching strain in feces | 5/9 | 2/10 | |
| Other adults[f] with clinical isolate–matching strain in feces | 0 | 0 | |
| Children with clinical isolate–matching strain in feces | 0 | 0 | |
| Pets with clinical isolate–matching strain in feces | 2/8 | 2/11 | |

[a]Households in which the veteran's index clinical *E. coli* isolate was fluoroquinolone-resistant.

[b]Households in which the veteran's index clinical *E. coli* isolate was fluoroquinolone-susceptible.

[c]*P* values shown if <.05. Comparisons involving dichotomous variables were assessed using the (n-1) chi-square test, whereas those involving continuous variables were assessed using the Mann-Whitney test.

[d]Unique-by-household strains, including fecal strains that matched the index clinical *E. coli* isolate (ie, the "matching clinical isolate" strains).

[e]Unique-by-household strains, excluding fecal strains that matched the index clinical *E. coli* isolate (ie, the "matching clinical isolate" strains).

[f]Nonveteran and nonspouse adults.

phylogroup F (Table 5). However, statistically significant associations were limited to phylogroups F (fecal predominance [negative]: $P = .02$) and B1 (household prevalence [negative]: $P < .001$) (Table 4).

Of the 15 studied clonal lineages and sublineages, 6 qualified numerically for statistical analysis, including (number of strains) ST95 (16), ST131 (14), ST131-*H*30 (13), ST73 (8), ST127 (8), and ST648 (5). Each such lineage except ST127 was associated significantly with ≥1 colonization outcome (Table 4). Specifically, ST131 and ST131-*H*30 predicted persistence (for each, $P = .002$) and household prevalence ($P = .04$ and $P = .02$, respectively), ST95 predicted sharing ($P = .02$), ST73 (negatively) predicted household prevalence ($P = .01$), and ST648 (negatively) predicted fecal predominance ($P = .02$).

Of the 49 studied virulence genes, 43 were detected; 38 qualified for statistical analysis. Of these, 18 (47%) were associated significantly with ≥1 colonization outcome, including fecal predominance (11 genes), persistence (5 genes), household prevalence (5 genes), and sharing (3 genes) (Table 4). Of the 18 colonization-associated genes, 1 was associated with 3 outcomes, 4 with 2 outcomes each, and 13 with 1 outcome each. Similarly, molecular ExPEC status (n = 61) predicted persistence ($P = .04$). Notably, the 11 fecal predominance–associated genes overlapped minimally (only *papAH* and *papC*) with

**Table 3. Relationship Between Host Demographic Subset and Colonization Outcomes[a] Among 139 Fecal *Escherichia coli* Strains**

| Host Subset[d,e] | Strains, No. | Persistence | | | | Sharing | | | | Prevalence | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Univar. | | Multivar.[a] | | Univar. | | Multivar.[b] | | Univar. | | Multivar.[c] | |
| | | OR[f] | *P* | OR | *P*[g] | OR | *P* | OR | *P* | Co | *P* | Co | *P* |
| Veteran | 63 | | | | | 1.88 | .02 | 2.93 | .02 | .31 | <.001 | .19 | .01 |
| Spouse | 50 | 2.63 | .01 | 2.63 | .02 | 4.56 | <.001 | 6.21 | <.001 | .32 | <.001 | .40 | <.001 |
| Child | 14 | | | | | | | | | .30 | .001 | | |

Abbreviations: Co., correlation; Multivar., multivariable; OR, odds ratio; Prevalence, household prevalence; Univar., univariable.

[a]Multivariable candidate predictor variables were spouse, matches CI, ST131-*H*30, *sat* (secreted auto-transporter toxin), *usp* (uropathogenic-specific protein), *kpsMII* (group 2 capsules), *ompT* (outer membrane protease), and ExPEC (extra-intestinal pathogenic *E. coli*, functionally defined as ≥2 of the following: *papAH* and/or *papC* [counted as 1: P fimbriae], *sfa/focDE* [S and F1C fimbriae], *afa/draBC* [Dr-binding adhesins], *iutA* [aerobactin receptor], and *kpsM*II).

[b]Multivariable candidate predictor variables were veteran, spouse, matches CI, ST95, *vat* (vacuolating toxin), and *kpsMII*.

[c]Multivariable candidate predictor variables were veteran, spouse, child, matches CI, phylogroup B1, ST73, ST131-*H*30, *papAH, iha* (adhesin-siderophore receptor), *usp,* and *kpsMII*.

[d]Host subset and outcome variables shown are those that yielded $P < .05$ for at least 1 (host subset vs outcome variable) comparison. No statistically significant associations were observed for the outcome variable "fecal predominance" or for host subsets: adult (nonspouse; n = 19) and pet (n = 19).

[e]Host subset refers to the subject from whom the fecal sample yielded the *Escherichia coli* strain. Some strains were present in multiple host subsets.

[f]Odds ratios and correlations are shown only for significant associations.

[g]*P* values are shown where $P < .05$. Comparisons involving dichotomous variables (persistence and sharing) were assessed using the (n-1) chi-square test or binomial logistic regression, whereas those involving continuous variables (predominance, household prevalence) were assessed using the Mann-Whitney test.

**Table 4. Relationship Between Molecular Traits and Colonization Outcomes Among 139 Fecal *Escherichia coli* Strains**

| | | | Predominance | | | | Persistence | | | | Sharing | | | | Prevalence | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Univar. | | Multivar. [a] | | Univar. | | Multivar. [b] | | Univar. | | Multivar. [c] | | Univar. | | Multivar. [d] | |
| Traits [e] | | Strains, No. | Cor. | *P* [f] | Co | *P* | OR [g] | *P* | OR | *P* | OR | *P* | OR | *P* | Co | *P* | Co | *P* |
| Matches CI | | 16 | | | | | 5.66 | <.001 | 4.20 | .02 | 3.24 | .005 | | | .37 | <.001 | .33 | <.001 |
| Phylogroup | B1 | 22 | | | | | | | | | | | | | –.35 | <.001 | | |
| | F | 7 | –.20 | .02 | –.17 | .04 | | | | | | | | | | | | |
| Lineages | ST131 | 14 | | | | | 1.36 | .002 | | | | | | | .18 | .04 | | |
| | ST131-*H*30 | 13 | | | | | 5.12 | .002 | | | | | | | .20 | .02 | | |
| | ST73 | 8 | | | | | | | | | | | | | –.22 | .01 | –.14 | .04 |
| | ST95 | 16 | | | | | | | | | 3.24 | .02 | | | | | | |
| | ST648 | 5 | –.19 | .02 | | | | | | | | | | | | | | |
| Genes [h] | | | | | | | | | | | | | | | | | | |
| Adhesin | *papAH* | 37 | .28 | .001 | | | | | | | | | | | .17 | .048 | | |
| | *papC* | 37 | .28 | .001 | | | | | | | | | | | .17 | .048 | | |
| | *papG* II | 25 | .22 | .01 | | | | | | | | | | | | | | |
| | *papG* | 39 | .17 | .04 | | | | | | | | | | | | | | |
| | *papEF* | 36 | .26 | .002 | | | | | | | | | | | | | | |
| | *sfaS* | 20 | .18 | .04 | | | | | | | | | | | | | | |
| | *iha* | 23 | | | | | 2.22 | .02 | | | | | | | .17 | .04 | .15 | .04 |
| Toxin | *hlyD* | 22 | .20 | .02 | .21 | .01 | | | | | | | | | | | | |
| | *hlyF* | 9 | .22 | .01 | .19 | .02 | | | | | | | | | | | | |
| | *sat* | 27 | | | | | 2.45 | .01 | | | | | | | | | | |
| | *vat* | 47 | | | | | | | | | 2.27 | .04 | 2.42 | .04 | | | | |
| Siderophore | *fyuA* | 78 | | | | | 1.67 | .04 | | | | | | | | | | |
| | *ireA* | 21 | .21 | .02 | | | | | | | | | | | | | | |
| Protectin | *iss* | 6 | .19 | .03 | | | | | | | | | | | | | | |
| | *kpsMII* | 70 | | | | | 2.59 | .004 | | | 2.08 | .04 | | | .25 | .004 | | |
| | *usp* | 59 | | | | | 1.69 | .02 | | | | | | | .21 | .02 | | |
| Mobility | H7 *fliC* | 18 | .22 | .01 | –.17 | .04 | | | | | | | | | | | | |
| Misc. | *ompT* | 87 | | | | | | | | | 1.82 | .02 | | | | | | |
| | ExPEC | 61 | | | | | 2.19 | .04 | | | | | | | | | | |

Abbreviations: CI, clinical isolate; Cor., correlation; ExPEC, extraintestinal pathogenic *E. coli;* Misc. miscellaneous; Multivar., multivariable; OR, odds ratio; Prevalence, household prevalence; Univar., univariable.

[a]Multivariable candidate predictor variables included phylogroup F, *papAH* (P fimbriae), *hlyF* (variant hemolysin), *sfaS* (S fimbriae adhesin), *hlyD* (alpha hemolysin), *ireA* (siderophore receptor), and H7 *fliC* (flagellar variant).

[b]Multivariable candidate predictor variables included spouse, matches CI, ST131-*H*30, *sat* (secreted auto-transporter toxin), *usp* (uropathogenic-specific protein), *kpsMII* (group 2 capsules), *ompT* (outer membrane protease), and ExPEC (defined as ≥ 2 of the following: *papAH* and/or *papC* [counted as 1: P fimbriae], *sfa/focDE* [S and F1C fimbriae], *afa/draBC* [Dr-binding adhesins], *iutA* [aerobactin receptor], and *kpsM*II).

[c]Multivariable candidate predictor variables included veteran, spouse, matches CI, ST95, *vat* (vacuolating toxin), and *kpsMII*.

[d]Multivariable candidate predictor variables were veteran, spouse, child, matches CI, phylogroup B1, ST73, ST131-*H*30, *papAH, iha* (adhesin-siderophore receptor), *usp,* and *kpsMII*.

[e]Molecular traits shown are those that yielded $P < .05$ for at least 1 outcome variable. Three categories of molecular traits were excluded from the table: (i) Traits detected in ≥5 strains but yielding $P > .05$: ST127, *papGIII* (P fimbriae), *sfa/focDE*, *focG* (F1C fimbriae adhesin), *hra* (heat-resistant agglutinin), *fimH* (type 1 fimbriae adhesin), *cnf1* (cytotoxic necrotizing factor), *cdtB* (cytolethal distending toxin), *pic* (protein associated with intestinal colonization), *iroN* (salmochelin receptor), *iutA*, *neuB* (K1 capsule variant), *kfliC* (K5 capsule variant), K2/K100 (capsule variants), *rfc* (O4 lipopolysaccharide synthesis), *cvaC* (colicin/microcin V), *traT* (serum resistance–associated), *ibeA* (invasion of brain endothelium), *malX* (pathogenicity island marker), *clbB* and *clbN* (colibactin polyketide synthesis). (ii) Traits detected in <5 strains (too few for statistical analysis): phylogroup C, ST12, ST141, ST372, ST14, ST405, 016, CGA, ST131-*H*30Rx, *papGI* (P fimbriae), *kpsMTIII* (group 3 capsules), *afa/draBC*, *tsh* (temperature-sensitive hemagglutinin), and *astA* (enteroaggregative *Escherichia coli* toxin). (iii) Traits sought but not detected: ST144, O15 clonal group (ST31 complex), *afaE8* (variant afimbrial adhesin), *bmaE* (M fimbriae), *gafD* (G fimbriae), *F17* (adhesin variant), *clpG* (adhesin variant), K15 *kpsM* (capsule variant).

[f]*P* values are shown where $P < .05$. For univariable analysis, comparisons involving dichotomous variables (persistence and sharing) were assessed using the (n-1) chi-square test or binomial logistic regression, whereas those involving continuous variables (predominance, household prevalence) were assessed using the Mann-Whitney test or regression. Multivariable analysis used backward stepwise variable entry.

[g]Odds ratios and correlations are shown only for significant associations ($P < .05$).

[h]Virulence genes: p*apAH, papC, papGII, papG*, and *papEF* (P fimbriae) *sfaS*, *iha*, *hlyD*, *hlyF*, *sat*, *vat*, *fyuA* (yersiniabactin receptor), *ireA*, *ompT*, *iss* (increased serum survival), *kpsMII*, *usp*, H7 *fliC*, and molecular ExPEC status.

**Table 5. Rank Order of Phylogroups for Colonization Outcomes Among 139 Fecal *Escherichia coli* Strains**

| | | Rank Order of Phylogroups for Each Outcome Variable[b] | | | |
|---|---|---|---|---|---|
| Phylogroup[a] | Strains, No. | Predominance | Persistence | Sharing | Prevalence |
| A | 29 | 3 | 5 | 4 | 3 |
| B1 | 23 | 4 | 2 | 5 | 5 |
| B2 | 67 | 2 | 1 | 2 | 2 |
| D | 12 | 1 | 2 | 3 | 1 |
| F | 7 | 5 | 5 | 1 | 4 |

Abbreviations: Prevalence, household prevalence.

[a]Phylogroups shown are those that accounted for at least 5 strains. Phylogroup C (not shown) accounted for only 1 strain. Phylogroup E did not account for any strains.

[b]Numerical rank order for the corresponding colonization outcome (1 = highest, 5 = lowest).

those that predicted within-household colonization patterns (Table 4).

### Colonization Outcomes: Antibiotic Resistance

Of the 24 resistance markers, 14 were detected in ≥5 fecal strains each, so they could be analyzed statistically. Of these, 7 (50%) corresponded significantly with ≥1 outcome variable in univariable analyses (Table 6). For persistence, the 4 significant predictors were piperacillin ($P \leq .001$), amoxicillin-clavulanic acid ($P = .02$), ciprofloxacin/levofloxacin ($P = .006$), and nalidixic acid ($P = .02$). For prevalence, the 5 significant predictors were piperacillin ($P \leq .001$), amoxicillin-clavulanic acid ($P = .001$), nalidixic acid ($P = .04$), ampicillin ($P = .02$), and ampicillin-sulbactam ($P = .046$). For sharing, the only significant predictor (negative) was streptomycin ($P = .03$). No antibiotic predicted fecal predominance.

### Multivariable Analysis

Given the numerous univariable host and molecular predictors of colonization outcomes, multivariable analysis was used to identify primary associations using a single representative for each cluster of correlated variables (correlation coefficient ≥ 0.7). In backward stepwise multivariable models, veteran source remained predictive of sharing and household prevalence ($P = .02$ and $P = .01$, respectively) and spouse source remained predictive of persistence, sharing, and household prevalence ($P = .02$, $P < .001$, and $P < .001$, respectively) (Table 3). Clinical isolate–matching (vs fecal only) status remained predictive of persistence and household prevalence ($P = .02$ and $P < .001$, respectively). One or more molecular traits predicted fecal predominance (phylogroup F: $P = .04$; *hlyF*: $P = .02$; *hlyD*: $P = .01$; H7 *fliC*: $P = .04$), sharing (vat: $P = .04$), and household prevalence (ST73: $P = .04$ [negative], *iha* [adhesin-siderophore receptor]: $P = .04$). By contrast, none predicted persistence (Table 4).

Additional multivariable analyses involving molecular traits only (ie, excluding demographic subset and clinical isolate–matching status) were used to further assess the relationship between molecular traits and colonization outcomes. These models identified slightly different predictive molecular traits than did the initial models: For persistence, ST131-*H*30 was newly predictive; for sharing, *vat* lost significance, whereas ST95 was newly predictive; and for household prevalence, *iha* lost significance, ST73 remained predictive, and phylogroup B1 was newly predictive (Supplementary Tables 3–6). Models using forced or forward stepwise variable entry yielded results largely similar to those of the backward stepwise entry models, although forced entry usually

**Table 6. Relationship Between Antibiotic Resistance and Colonization Outcome**

| | | Persistence | | Sharing | | Prevalence | |
|---|---|---|---|---|---|---|---|
| Antibiotic[a] | No. | Odds Ratio[b] | *P* | Odds Ratio | *P* | Correlation | *P* |
| Ampicillin | 50 | - | - | - | - | .21 | .02 |
| Piperacillin | 28 | 3.63 | <.001 | | | .34 | <.001 |
| Amoxicillin-clavulanic acid | 35 | 1.87 | .02 | - | - | .30 | .001 |
| Ampicillin-sulbactam | 23 | | | | | .17 | .05 |
| Ciprofloxacin/levofloxacin | 24 | 2.76 | .006 | - | - | - | - |
| Naladixic acid | 29 | 1.65 | .02 | | | .17 | .04 |
| Streptomycin | 36 | - | - | .09 | .03 | - | - |

[a]Antibiotics and outcome variables shown are those that yielded $P < .05$ for at least 1 (antibiotic-vs-outcome variable) comparison. No statistically significant associations were observed for the outcome variable "fecal predominance" or for the remaining 17 antibiotics (amikacin, aztreonam, cefazolin, cefotaxime, cefoxitin, ceftazidime, ceftriaxone, cephalothin chloramphenicol, ertapenem, fosfomycin, gentamicin, imipenem, piperacillin-tazobactam, teicoplanin, trimethoprim, trimethoprim-sulfamethoxazole).

[b]All associations were outcomes of univariable analysis. Notably, in an exploratory multivariable analysis including antibiotics along with all other strain characteristic variables (ie, host demographic and molecular traits), streptomycin was the only resistance variable that significantly predicted any colonization outcome (negatively predictive of household sharing: not shown).

identified fewer predictors than did the stepwise approaches (Supplementary Tables 3–6).

## DISCUSSION

Our longitudinal surveillance of fecal *E. coli* strains from 22 veterans with a clinical *E. coli* isolate and their 46 human and animal household members yielded 3 main findings, with potentially important implications regarding the ecology of this important extraintestinal pathogen. Specifically, within-household fecal colonization behaviors were associated with (i) multiple virulence-associated clonal lineages and genes, (ii) host demographic subset, and (iii) matching the household's index clinical isolate. This suggests that traits associated traditionally with virulence likely also contribute to fecal colonization behavior.

Our first main finding was that extraintestinal virulence-associated lineages and genes predicted fecal colonization behavior. The observed colonization behaviors of ST131 and its *H*30 subclone are consistent with prior reports of ST131's remarkable ability to disseminate and persist [7, 11, 13, 15, 17, 18, 30]. Likewise, here ST95, a prominent endemic ExPEC lineage, also exhibited increased within-household strain sharing. Conceivably, superior colonization/sharing ability contributes to these lineages' epidemiological success. By contrast, ST73, another prominent cause of extraintestinal *E. coli* infections [31] that in previous studies exhibited increased sharing and persistence [12, 32], here was associated negatively with household prevalence. These between-study and between-lineage differences suggest that although virulence and commensalism may align generally, exceptions occur, perhaps related to specific lineages and/or host populations or stochastic factors.

Additionally, 18 virulence genes and molecularly defined ExPEC status predicted 1 or more colonization outcomes. These genes represented diverse functional categories, precluding simple mechanistic interpretations and suggesting instead a multifactorial effect.

Of the 18 colonization-associated genes, 8 (*iha, hlyF, sat, vat, ireA, ompT, iss,* and H7 *fliC*) were newly identified as predicting colonization outcomes, whereas 10 had been so identified previously. Of these 10, 5 represent the *pap* (P fimbriae) operon, which previously was associated with gut colonization and/or strain sharing among men with febrile urinary tract infection and their sex partners [12], infants and schoolgirls [9, 33], members of a 3-subject household [34], and women with cystitis and their sex partners [8]. The other 5 previously identified genes (*fyuA*, *kpsMII*, *sfaS*, *hlyD*, and *usp* [uropathogenic-specific protein]) likewise were associated previously with colonization and/or strain sharing in diverse contexts and hosts [12, 34–36].

By contrast, the 8 newly identified colonization-associated genes, which similarly represented diverse functional categories, were assessed previously in only 1 or 2 colonization studies each [12, 19]. Our study's larger number of households and more extended follow-up may underlie our novel identification of colonization-related associations for these genes. Collectively, the present findings confirm and extend previous evidence that so-called extraintestinal "virulence factors," or traits linked to them, may contribute to gut colonization and/or strain sharing.

Here, phylogroups B2 and D ranked highest for fecal predominance, persistence, and prevalence and second highest (after phylogroup F) for sharing. This aligns generally with evidence that phylogroup B2—the most extraintestinal pathogenic *E. coli* phylogroup—corresponds with within-household strain sharing [12, 14, 19, 32, 37, 38]. Likewise, here phylogroup B1, generally regarded as less virulent, was associated negatively with fecal colonization. Paradoxically, phylogroup F, a recently delineated phylogroup associated with antimicrobial resistance and virulence [39, 40], behaved similarly to group B1. Collectively, these phylogroup-related associations provide added support for a linkage between colonization and extraintestinal virulence.

Our second main finding was that even in multivariable models 1 or more host demographic subsets predicted persistence (spouse), strain sharing (spouse and veteran), and overall household prevalence (spouse and veteran) but not fecal predominance. Prior studies have associated *E. coli* strain sharing with sexual behaviors [3, 8, 12], which we did not assess; sexual or other types of host–host contact may underlie our observed associations of strain sharing with spouses and veterans. Our findings regarding demographic subsets suggest that diverse factors other than inherent strain characteristics, including host biology and behaviors, also influence *E. coli* colonization phenotypes.

Our third main finding was that clinical isolate–matching fecal strains exhibited increased household colonization behavior, specifically sharing, persistence, and overall household prevalence. This additionally supports a link between commensalism and virulence. Notably, clinical isolate–matching fecal strains were mostly (10/16) fluoroquinolone-resistant, and 6/10 such strains were ST131-*H*30, underscoring the clinical importance of colonizing ExPEC strains.

Some propose that *E. coli* virulence genes are byproducts of commensalism, with certain traits promoting both colonization and infection-causing ability [6], as supported by human epidemiological observations and animal model studies [7–9, 12–14, 19, 21, 32, 41]. Our work supports this paradigm of co-evolution of virulence and commensalism, expands the number of clonal lineages and virulence factors known to be associated with gut colonization, and suggests that host factors also may play an important role in *E. coli* household colonization dynamics.

In addition to host and molecular factors, we found that resistance to several common antibiotics predicted increased persistence of strains in serial fecal samples (piperacillin, amoxicillin-clavulanic acid, ciprofloxacin/levofloxacin,

nalidixic acid) and overall prevalence in household fecal samples (piperacillin, ampicillin, amoxicillin-clavulanic acid, ampicillin-sulbactam, nalidixic acid). By contrast, antibiotic resistance predicted neither increased predominance of a strain within individual fecal samples nor increased sharing among household members.

Our study has limitations. First, the large number of candidate predictor variables, although appropriate for an exploratory study, risked finding associations by chance alone, whereas small numbers in certain subgroups risked missing true associations. Second, the study's observational nature precluded causality assessments. Third, use of ciprofloxacin-supplemented plates may have biased toward detection of ciprofloxacin-resistant strains, although only 6 (4% of 139) strains came only from ciprofloxacin-supplemented plates. Fourth, the focus on veterans with a clinical *E. coli* isolate may have biased toward detecting associations involving clinical isolate–matching strains. Fifth, we studied only a fraction of all possible epidemiological and bacterial variables that might influence colonization behavior, for example, host comorbidities, exposures (including antibiotic use), and behaviors, and other genes, gene expression/regulation, and the gut microbiota and phageome.

Our study also has strengths. First, it uniquely assessed virulence genes and phylogenetic subsets as colonization predictors within >20 multimember households. Second, the extensive household fecal sampling permitted robust analyses of sharing and persistence of fecal *E. coli* strains. Third, use of generic index clinical *E. coli* isolates provided clinical relevance, unbiased by focusing on dramatic clinical scenarios [18]. Fourth, in contrast to prior studies involving children [21, 30, 33, 36, 42], our study involved mostly adults (45/49 [92%] of human subjects)—making its results more generalizable to adults and to veterans, who usually are elderly men [43], an understudied population.

In summary, multiple host and bacterial characteristics predicted *E. coli* fecal colonization behaviors among veterans and their household members. These findings support and extend the concept that recognized pathogenic lineages (eg, ST131) and putative extraintestinal virulence factors (eg, P fimbriae, other adhesins, toxins, siderophores, and protectins) may feature importantly in both individual- and household-level gut colonization. Thus, many "virulence factors" may also be "colonization factors," and many "virulent lineages" also "elite colonizing lineages." Additionally, host demographic subset strongly predicted household colonization; thus, host biology and/or behaviors likely contribute importantly to *E. coli* colonization patterns. Lastly, clinical isolate–matching fecal strains colonized most extensively, underscoring the relationship between superior gut colonization and infection-causing proficiency. Given the important clinical implications of *E. coli* colonization, our findings suggest that future interventions that modulate colonization behavior might represent new ways of preventing human disease.

## 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

We thank Brian D. Johnston for technical assistance.

***Financial support.*** This work was supported by the Office of Research and Development, Department of Veterans Affairs (grant numbers 1 I01 CX000192-01, 1 I01 CX000920-01A1, and 2 I01 CX000920-04 to J.R.J.); Intramural Bridge Grant, Medical School and University Medical Foundation (University of Minnesota to J.R.J.); and the National Institute of Allergy and Infectious Diseases of the National Institute of Health (Antibacterial Resistance Leadership Group, award number UM1AI104681 to J.R.J.).

***Potential conflicts of interest.*** James R. Johnson has received grants or consultancies from Achaogen/Cipla, Allergan, Janssen/Crucell, Melinta, Merck, Shionogi, Syntiron, and Tetraphase and has a patent application for tests to detect specific *E. coli* strains. The other authors report no financial 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.

***Prior presentation.*** This work was presented in part at IDWeek 2013 and IDWeek 2018 (posters).

## References

1. Grüneberg RN. Relationship of infecting urinary organism to the faecal flora in patients with symptomatic urinary infection. Lancet **1969**; 2:766–8.
2. Yamamoto S, Tsukamoto T, Terai A, et al. Genetic evidence supporting the fecal-perineal-urethral hypothesis in cystitis caused by *Escherichia coli*. J Urol **1997**; 157:1127–9.
3. Johnson JR, Owens K, Gajewski A, Clabots C. *Escherichia coli* colonization patterns among human household members and pets, with attention to acute urinary tract infection. J Infect Dis **2008**; 197:218–24.
4. Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, et al. Sequential acquisition of virulence and fluoroquinolone resistance has shaped the evolution of *Escherichia coli* ST131. mBio **2016**; 7:e00347–16.
5. Nicolas-Chanoine MH, Bertrand X, Madec JY. *Escherichia coli* ST131, an intriguing clonal group. Clin Microbiol Rev **2014**; 27:543–74.
6. Le Gall T, Clermont O, Gouriou S, et al. Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. Mol Biol Evol **2007**; 24:2373–84.
7. Madigan T, Johnson JR, Clabots C, et al. Extensive household outbreak of urinary tract infection and intestinal colonization due to extended-spectrum β-lactamase-producing *Escherichia coli* sequence type 131. Clin Infect Dis **2015**; 61:e5–12.
8. Foxman B, Manning SD, Tallman P, et al. Uropathogenic *Escherichia coli* are more likely than commensal *E. coli* to be shared between heterosexual sex partners. Am J Epidemiol **2002**; 156:1133–40.
9. Nowrouzian F, Wold AE, Adlerberth I. P fimbriae and aerobactin as intestinal colonization factors for *Escherichia coli* in Pakistani infants. Epidemiol Infect **2001**; 126:19–23.
10. Adlerberth I, Hanson LA, Svanborg C, et al. Adhesins of *Escherichia coli* associated with extra-intestinal pathogenicity confer binding to colonic epithelial cells. Microb Pathog **1995**; 18:373–85.
11. Johnson JR, Anderson JT, Clabots C, et al. Within-household sharing of a fluoroquinolone-resistant *Escherichia coli* sequence type ST131 strain causing pediatric osteoarticular infection. Pediatr Infect Dis J **2010**; 29:473–5.
12. Ulleryd P, Sandberg T, Scheutz F, et al. Colonization with *Escherichia coli* strains among female sex partners of men with febrile urinary tract infection. J Clin Microbiol **2015**; 53:1947–50.
13. Overdevest I, Haverkate M, Veenemans J, et al. Prolonged colonisation with *Escherichia coli* O25:ST131 versus other extended-spectrum

beta-lactamase-producing *E. coli* in a long-term care facility with high endemic level of rectal colonisation, the Netherlands, 2013 to 2014. Euro Surveill **2016**; 21:42.
14. Moreno E, Andreu A, Pigrau C, et al. Relationship between *Escherichia coli* strains causing acute cystitis in women and the fecal *E. coli* population of the host. J Clin Microbiol **2008**; 46:2529–34.
15. Ludden C, Cormican M, Vellinga A, et al. Colonisation with ESBL-producing and carbapenemase-producing Enterobacteriaceae, vancomycin-resistant enterococci, and meticillin-resistant *Staphylococcus aureus* in a long-term care facility over one year. BMC Infect Dis **2015**; 15:168.
16. Guo S, Brouwers HJ, Cobbold RN, et al. Fluoroquinolone-resistant extraintestinal pathogenic *Escherichia coli*, including O25b-ST131, isolated from faeces of hospitalized dogs in an Australian veterinary referral centre. J Antimicrob Chemother **2013**; 68:1025–31.
17. Johnson JR, Miller S, Johnston B, et al. Sharing of *Escherichia coli* sequence type ST131 and other multidrug-resistant and urovirulent *E. coli* strains among dogs and cats within a household. J Clin Microbiol **2009**; 47:3721–5.
18. Johnson JR, Davis G, Clabots C, et al. Household clustering of *Escherichia coli* sequence type 131 clinical and fecal isolates according to whole genome sequence analysis. Open Forum Infect Dis **2016**; 3:ofw129.
19. Johnson JR, Clabots C, Kuskowski MA. Multiple-host sharing, long-term persistence, and virulence of *Escherichia coli* clones from human and animal household members. J Clin Microbiol **2008**; 46:4078–82.
20. Johnson JR, Russo TA. Molecular epidemiology of extraintestinal pathogenic *Escherichia coli*. EcoSal Plus **2018**; 8:ESP-0004-2017.
21. Gurnee EA, Ndao IM, Johnson JR, et al. Gut colonization of healthy children and their mothers with pathogenic ciprofloxacin-resistant *Escherichia coli*. J Infect Dis **2015**; 212:1862–8.
22. Brikun I, Suziedelis K, Berg DE. DNA sequence divergence among derivatives of *Escherichia coli* K-12 detected by arbitrary primer PCR (random amplified polymorphic DNA) fingerprinting. J Bacteriol **1994**; 176:1673–82.
23. Mohamed M, Clabots C, Porter SB, et al. Large fecal reservoir of *Escherichia coli* sequence type 131-H30 subclone strains that are shared within households and resemble clinical ST131-H30 isolates. J Infect Dis **2020**; 221:1659–68.
24. Johnson JR, Nicolas-Chanoine MH, DebRoy C, et al. Comparison of *Escherichia coli* ST131 pulsotypes, by epidemiologic traits, 1967–2009. Emerg Infect Dis **2012**; 18:598–607.
25. Clermont O, Christenson JK, Denamur E, Gordon DM. The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. Environ Microbiol Rep **2013**; 5:58–65.
26. Johnson JR, Johnston BD, Porter SB, et al. Rapid emergence, subsidence, and molecular detection of *Escherichia coli* sequence type 1193-fimH64, a new disseminated multidrug-resistant commensal and extraintestinal pathogen. J Clin Microbiol **2019**; 57.
27. Johnson JR, Murray AC, Gajewski A, et al. Isolation and molecular characterization of nalidixic acid-resistant extraintestinal pathogenic *Escherichia coli* from retail chicken products. Antimicrob Agents Chemother **2003**; 47:2161–8.
28. Clermont O, Johnson JR, Menard M, Denamur E. Determination of *Escherichia coli* O types by allele-specific polymerase chain reaction: application to the O types involved in human septicemia. Diagn Microbiol Infect Dis **2007**; 57:129–36.
29. Campbell I. Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. Stat Med **2007**; 26:3661–75.
30. Huang IF, Lee WY, Wang JL, et al. Fecal carriage of multidrug-resistant *Escherichia coli* by community children in Southern Taiwan. BMC Gastroenterol **2018**; 18:86.
31. Riley LW. Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. Clin Microbiol Infect **2014**; 20:380–90.
32. Nowrouzian FL, Clermont O, Edin M, et al. *Escherichia coli* phylogenetic B2 subgroups in the infant gut microbiota: predominance of uropathogenic lineages in Swedish infants and enteropathogenic lineages in Pakistani infants. Appl Environ Microbiol **2019**; 85:e01681-19.
33. Nowrouzian F, Adlerberth I, Wold AE. P fimbriae, capsule and aerobactin characterize colonic resident *Escherichia coli*. Epidemiol Infect **2001**; 126:11–8.
34. Murray AC, Kuskowski MA, Johnson JR. Virulence factors predict *Escherichia coli* colonization patterns among human and animal household members. Ann Intern Med **2004**; 140:848–9.
35. Moreno E, Johnson JR, Pérez T, et al. Structure and urovirulence characteristics of the fecal *Escherichia coli* population among healthy women. Microbes Infect **2009**; 11:274–80.
36. Ostblom A, Adlerberth I, Wold AE, Nowrouzian FL. Pathogenicity island markers, virulence determinants malX and usp, and the capacity of *Escherichia coli* to persist in infants' commensal microbiotas. Appl Environ Microbiol **2011**; 77:2303–8.
37. Nowrouzian FL, Adlerberth I, Wold AE. Enhanced persistence in the colonic microbiota of *Escherichia coli* strains belonging to phylogenetic group B2: role of virulence factors and adherence to colonic cells. Microbes Infect **2006**; 8:834–40.
38. Nowrouzian FL, Wold AE, Adlerberth I. *Escherichia coli* strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. J Infect Dis **2005**; 191:1078–83.
39. Pitout JD. Extraintestinal pathogenic *Escherichia coli*: a combination of virulence with antibiotic resistance. Front Microbiol **2012**; 3:9.
40. Vangchhia B, Abraham S, Bell JM, et al. Phylogenetic diversity, antimicrobial susceptibility and virulence characteristics of phylogroup F *Escherichia coli* in Australia. Microbiology (Reading) **2016**; 162:1904–12.
41. Diard M, Garry L, Selva M, et al. Pathogenicity-associated islands in extraintestinal pathogenic *Escherichia coli* are fitness elements involved in intestinal colonization. J Bacteriol **2010**; 192:4885–93.
42. Nowrouzian F, Hesselmar B, Saalman R, et al. *Escherichia coli* in infants' intestinal microflora: colonization rate, strain turnover, and virulence gene carriage. Pediatr Res **2003**; 54:8–14.
43. United States Census Bureau. State veteran data. Available at: https://www.census.gov/topics/population/veterans.html. Accessed 22 November 2019.