The Journal of Infectious Diseases

MAJOR ARTICLE

Accessory Traits and Phylogenetic Background Predict *Escherichia coli* Extraintestinal Virulence Better Than Does Ecological Source

James R. Johnson,[1,2] Brian D. Johnston,[1,2] Stephen Porter,[1,2] Paul Thuras,[3] Maliha Aziz,[3] and Lance B. Price[3]

[1]Veterans Affairs Medical Center and [2]University of Minnesota, Minneapolis, Minnesota; and [3]George Washington University, Washington, D. C.

Background. The distinguishing characteristics of extraintestinal pathogenic *Escherichia coli* (ExPEC) strains are incompletely defined.

Methods. We characterized 292 diverse-source human *Escherichia coli* isolates (116 from fecal specimens, 79 from urine specimens [of which 39 were from patients with cystitis and 40 were from patients with pyelonephritis], and 97 from blood specimens) for phylogenetic group, sequence type complex (STc), and 49 putative extraintestinal pathogenic *E. coli* (ExPEC)–associated virulence genes. We then assessed these traits and ecological source as predictors of illness severity in a murine sepsis model.

Results. The study isolates exhibited a broad range of virulence in mice. Most of the studied bacterial characteristics corresponded significantly with experimental virulence, as did ecological source and established molecular definitions of ExPEC and uropathogenic *E. coli* (UPEC). Multivariable modeling identified the following bacterial traits as independent predictors of illness severity both overall and among the fecal and clinical (ie, urine and blood) isolates separately: *fyuA* (yersiniabactin receptor), *kpsM* K1 (K1 capsule), and *kpsM* II (group 2 capsules). Molecular UPEC status predicted virulence independently only among fecal isolates. Neither ecological source (ie, clinical vs fecal) nor molecular ExPEC status added predictive power to these traits, which accounted collectively for up to 49% of the observed variation in virulence.

Conclusions. Among human-source *E. coli* isolates, specific accessory traits and phylogenetic/clonal backgrounds predict experimental virulence in a murine sepsis model better than does ecological source.

Keywords. *Escherichia coli*; virulence; sepsis; mouse models; phylogenetic groups; virulence factors; sequence types; clinical isolates; fecal isolates.

Extraintestinal *Escherichia coli* infections cause considerable morbidity and mortality and increased healthcare costs [1]. Most such infections are due to distinctive *E. coli* strains, termed extraintestinal pathogenic *E. coli* (ExPEC) [2] or uropathogenic *E. coli* (UPEC) on the basis of their enhanced ability to cause extraintestinal disease, including urinary tract infection [3]. However, the distinguishing characteristics of ExPEC or UPEC strains remain incompletely defined.

Molecular epidemiological comparisons of isolates from different ecological sources (eg, fecal vs clinical isolates) can be informative [3–5]. However, they risk confounding by host compromise, which allows low-virulence strains to cause disease [6, 7], and by the intestinal reservoir of ExPEC [8], which creates an ExPEC subset among fecal surveillance isolates [9, 10]. By contrast, animal challenge studies, despite their limitations, provide a direct readout of intrinsic extraintestinal virulence, allowing nonconfounded comparisons of bacterial traits with virulence.

Previous studies used this approach to study collections of isolates and identified various *E. coli* phylogenetic subsets and accessory traits (ie, putative virulence factors) as statistical correlates of virulence in diverse animal models [3, 6, 11–14]. However, these studies assessed a limited number and source diversity of isolates and range of bacterial traits. Accordingly, we sought to identify bacterial correlates of experimental virulence in a murine sepsis model by using a large set of extensively characterized *E. coli* isolates from diverse ecological contexts, locales, periods, and host populations.

METHODS

Isolates

The 292 *E. coli* study isolates were selected from multiple published collections [5, 6, 15–17], with a target number of approximately 20 (or 40, for the veterans fecal collection) presumptive ExPEC and non-ExPEC isolates per collection, as available (Table 1), for 300 total isolates. Collections were chosen to give a broad distribution by year of isolation (1981 through

Received 14 December 2017; editorial decision 16 July 2018; accepted 31 July 2018; published online August 2, 2018.

Correspondence: J. R. Johnson, MD, Infectious Diseases 111F, VA Medical Center, 1 Veterans Dr, Minneapolis, MN 55417 (johns007@umn.edu).

The Journal of Infectious Diseases® 2019;219:121–32

Published by Oxford University Press for the Infectious Diseases Society of America 2018. This work is written by (a) US Government employee(s) and is in the public domain in the US.

DOI: 10.1093/infdis/jiy459

2000), ecological source (ie, surveillance fecal specimens and clinical specimens [urine specimens from patients with cystitis or pyelonephritis and blood specimens from patients with urosepsis or bacteremia]), host population (ie, male inpatients at Veterans Affairs [VA] medical centers and ambulatory and hospitalized women), and presumptive ExPEC status (based on established molecular criteria) [18]. For the fecal [5, 15], cystitis [15], and pyelonephritis [16] collections, which contained abundant presumptive ExPEC and non-ExPEC isolates, similar numbers of ExPEC and non-ExPEC isolates were selected randomly. For the VA bacteremia collection [5], presumptive ExPEC isolates, which predominated, were selected randomly, whereas all 9 presumptive non-ExPEC isolates were used. For the Seattle urosepsis collection [17], for which experimental virulence data were available [6], all isolates were used, irrespective of presumptive ExPEC status. Of the 300 initially selected isolates, 8 were excluded for technical reasons, leaving 292 isolates as the final study population (Table 1).

Genome Sequencing

Genomes were sequenced and analyzed as described elsewhere [9]. Pooled paired-end libraries were sequenced on an Illumina MiSeq and Genome Analyzer IIx to a read length of ≥100 base pairs, at a mean coverage depth (±SD) of 58.27-fold ± 35.4-fold. After alignment of short-read sequences to a reference genome by using BWA-mem (v.0.7.12) [19], single-nucleotide polymorphisms (SNPs) were called using GATK (v.3.5) [20], recombinant regions were identified and removed using Gubbins (v.2.1) [21], and the resultant SNP matrix was used to construct phylogenetic trees in PhyML with Smart Model Selection [22].

Sequence Types (STs) and Phylogenetic Groups

We determined STs by extracting from the genomes the 7 ST-defining housekeeping loci used in the Achtman MLST system. By using Enterobase (available at: https://enterobase.warwick.ac.uk/), each sequence variant was assigned an allele designation, and each allele combination was assigned to an ST. STs were grouped by ST complex (STc) according to Enterobase or if they differed by 1 locus.

Phylogroups were provisionally assigned using the updated polymerase chain reaction (PCR)–based method of Clermont et al [23] and were definitively assigned on the basis of each isolate's placement within the phylogram. Outlier isolates were classified as having an undetermined phylogroup.

Virulence Genotyping

Forty-nine putative or proven extraintestinal virulence genes were sought by either PCR analysis (n = 47) or in silico analysis (n = 2). PCR was done using established protocols [6], duplicate boiled lysates as template DNA, and inclusion of positive and negative controls. For *yfcV* and *chuA*, genomes were screened using BLAST analysis, based on 90% similarity to reference sequences [9]. Based on previous epidemiological and experimental validation, isolates were classified as $ExPEC_{JJ}$ (per the criteria of J. Johnson) if positive for ≥2 of *papAH* and/or *papC* (P fimbriae), *sfa/focDE* (S and F1C fimbriae), *afa/draBC* (Dr-binding adhesins), *iutA* (aerobactin siderophore system), and *kpsM* II (group 2 capsules) [18]; and as $UPEC_{HM}$ (per the criteria of H. Mobley) if positive for ≥2 of *chuA* (heme uptake), *fyuA* (yersiniabactin siderophore system), *vat* (vacuolating toxin), and *yfcV* (adhesin) [3].

Table 1. Origins of the 292 *Escherichia coli* Study Isolates

Source (No. [%]), Syndrome	Context	Locale	Years	Isolates Selected From Source Collection, No.			$ExPEC_{JJ}$[a] Isolates in Source Collection, %	Reference(s)
				Total	$ExPEC_{JJ}$[a]	Non-$ExPEC_{JJ}$[a]		
Feces (116 [40])								
Not applicable[b]	Student health service[c]	Minneapolis, MN	1999–2000	39	22	17	37	[15]
Not applicable[b]	Hospitalized veterans[d]	Minneapolis, MN	1996–1999	77	41	36	38	[5]
Urine (79 [27])								
Cystitis	Student health service[c]	Minneapolis, MN	1999–2000	39	23	16	41	[15]
Pyelonephritis	Ambulatory women	Multicenter (USA)	1994–1997	40	21	19	69	[15, 16]
Blood (97 [33])								
Urosepsis	Four hospitals[e]	Seattle, WA	1981–1985	67	59	8	88	[6, 17]
Bacteremia	Hospitalized veterans[d]	Minneapolis, MN	1996–1999	30	21	9	81	[5]

[a]Isolates were defined as extraintestinal pathogenic *E. coli* ($ExPEC_{JJ}$) if ≥2 of the following virulence factor genes were present: *papAH* and/or *papC*, *sfa/focDE*, *afa/draBC*, *iutA*, and *kpsM* II.

[b]Donors of fecal specimens had no clinical evidence of infection.

[c]University of Minnesota.

[d]Minneapolis Veterans Affairs Medical Center.

[f]Harborview Medical Center, Public Health Service Hospital, Seattle Veterans Affairs Medical Center, and University of Washington Medical Center.

Murine Sepsis Model

An established murine subcutaneous sepsis model was used [6, 11, 14, 24]. The protocol was reviewed and approved by the local institutional animal care and use committee. Approximately 3×10^8 colony-forming units of bacteria in the exponential phase of growth were injected subcutaneously into female outbred Swiss-Webster mice (Harlan; Indianapolis, IN), using 5 mice initially and, if the initial standard error exceeded 20%, 5–10 additional mice. In parallel, laboratory strain MG1655 and pyelonephritis isolate CFT073 were injected into 5 mice each as negative and positive controls, respectively [11, 14, 24].

After inoculation, mice were observed for 72 hours and were scored daily for illness severity by a single experienced observer blinded to strain identity, using a 5-point scale (with a score of 1 denoting healthy; 2, mildly ill; 3, moderately ill; 4, severely ill/moribund; and 5, dead), with standardized criteria for each stage. Mice that died received a score of 5 for any remaining study days. Mice were euthanized on reaching stage 4 or surviving 72 hours. For mice challenged with a given strain, the mean of the daily illness scores was the strain's overall illness severity score, and the proportion of mice that died or reached stage 4 was the strain's lethality score.

Statistical Methods

Statistical testing of dichotomous variables was limited to those with an overall prevalence of 5%–95%. Comparisons of proportions were tested using the Fisher exact test or a χ^2 test, as appropriate. Comparisons involving continuous variables were tested using a 2-tailed *t* test. Because the virulence indicators were dichotomous and the outcomes were continuous, we used Spearman rank correlations to assess the strength of the associations between the potential predictors and outcomes. Concurrent assessment of multiple variables as predictors of a continuous dependent variable was done using multiple regression analysis, with both forced entry and forward and backward stepwise entry.

Because 5–15 mice were tested per strain, analysis at the mouse level conceivably could increase sample size and, thereby, statistical power but also could be confounded by clustering at the strain level. To determine whether outcomes could be analyzed validly at the mouse level and how much additional statistical power this would provide, we assessed the impact of clustering by strain, using an unconditional multilevel model, with severity as the dependent variable. This showed that strain-level effects accounted for 83% of the overall variance in illness severity, evidence that clustering by strain was quite significant and would therefore require adjustment in a by-mouse analysis, and that the increment in statistical power from a by-mouse analysis, once adjusted for by-strain clustering, would be small. Because of this and because adjustment for by-strain clustering would preclude stepwise multivariable modeling, we analyzed the data at the by-strain level.

For multivariable modeling, variables were selected as candidate predictors, using a multistage approach, with different partitions of the data set (ie, total and partial [fecal vs clinical isolates]), to assess consistency of results across source groups. First, within a given data set, all candidate predictor variables were assessed for their strength of correlation with the selected outcome (Table 2), and those with a correlation coefficient of ≥0.30 were assessed further. Next, we examined for intercorrelations among these predictors, and, where 2 predictors had a coefficient of >0.70, we chose the predictor most highly correlated with the outcome as the primary predictor variable. If 2 intercorrelated predictors were correlated equivalently with the outcome, the one that dominated more alternate intercorrelated predictors was used. In the end, all variables for entry into the regression models had coefficients of ≥0.30 for severity and <0.70 for each other. The same selection process was followed for each population partition and was done both with and without considering the composite variables, $UPEC_{HM}$ and $ExPEC_{JJ}$, as candidate predictors.

After selecting sets of candidate predictors, multiple regression models (forced and stepwise entry) were constructed for each population partition. For stepwise models, the criteria for entry and removal were *P* values of <.01 and <.05, respectively. Bootstrapping (with 500 iterations) was used to assess model stability.

RESULTS

Study Population Characteristics

The 292 study isolates represented diverse ecological sources, clinical contexts, host populations, locales, and periods (Table 1). They were predominantly from phylogroup B2 (167 [57%]), with minor contributions from groups D (41 [14%)], A (30 [10%]), B1 (30 [10%]), C (10 [3.4%]), F (9 [3%]), and E (2 [0.7%]); 3 (1%) had an undefined phylogroup. The 5 STc with a ≥5% prevalence were STc73 (43 [15%]), STc95 (37 [13%]), STc10 (17 [6%]), STc14 (17 [6%]), and STc69 (15 [5%]). Each of the 49 virulence genes sought except *clpG* (a non-P adhesin; not detected) was identified in from 0.3% (*papG* allele I and F17 adhesin) to 99% (*fimH* type 1 fimbria) of isolates. Overall, 186 isolates (64%) qualified molecularly as $ExPEC_{JJ}$, and 216 (74%) qualified as $UPEC_{HM}$; these variables were closely correlated (Spearman $\rho = 0.71$; $P < .001$).

Murine Sepsis Model Outcomes

In the murine sepsis model, mean illness severity over the 3-day observation period was distributed fairly evenly across the population (Figure 1). By contrast, the lethality percentage was strongly bimodal, with peaks at 0% and 100%. Mean illness severity and lethality percentage were highly correlated (Spearman $\rho = .92$; $P < .001$). For maximal power, illness severity was selected as the representative virulence outcome.

Table 2. Univariable Correlates of Experimental Virulence in a Murine Sepsis Model

	All Isolates (n = 292)					Fecal Isolates (n = 116)					Clinical Isolates (n = 176)				
		Mean Illness Severity Score					Mean Illness Severity Score					Mean Illness Severity Score			
Category, Trait[a]	No. (%)	Trait Absent	Trait Present	Correlation[b]	*P*[b]	No. (%)	Trait Absent	Trait Present	Correlation[b]	*P*[b]	No. (%)	Trait Absent	Trait Present	Correlation[b]	*P*[b]
Phylogenetic group															
Group A	30 (10)	3.6	2.6	–0.19	.001	14 (12)	3.4	2.7	–0.08	.26	16 (9)	3.5	2.5	–0.25	.002
Group B1	30 (10)	3.4	2.1	–0.31	<.001	22 (19)	3.4	2.0	–0.40	<.001	8 (5)	3.5	2.3	–0.19	.03
Group B2	167 (57)	2.6	3.8	0.46	<.001	61 (53)	2.4	3.8	0.52	<.001	106 (60)	2.8	3.8	0.40	<.001
Group D	41 (14)	3.3	3.0	–0.10	.12	15 (13)	3.2	2.6	–0.16	.06	26 (15)	3.4	3.2	–0.06	.47
STc															
STc10	17 (6)	3.3	2.8	–0.09	.12	6 (5)	3.1	3.2	0.03	.85	11 (6)	3.5	2.7	–0.17	.046
STc14	17 (6)	3.3	2.8	–0.09	.07	5 (4)	3.1	2.8	–0.03	.61	12 (7)	.35	2.8	–0.14	.06
STc69	15 (5)	3.3	3.0	–0.05	.41	6 (5)	3.1	2.4	–0.152	.25	9 (5)	3.4	3.4	–0.02	1.0
STc73	43 (15)	3.2	3.9	0.20	.001	4 (3)	3.0	4.4	0.27	<.001	34 (19)	3.3	3.7	0.13	.03
STc95	37 (13)	3.1	4.1	0.29	<.001	19 (16)	2.9	4.2	0.36	<.001	18 (10)	3.3	4.3	0.25	<.001
Adhesin															
papAH	145 (50)	2.8	3.8	0.37	<.001	42 (36)	2.7	3.7	0.35	<.001	103 (59)	2.9	3.8	0.35	<.001
papG II	114 (39)	3.0	3.7	0.29	<.001	29 (25)	2.9	3.8	0.28	.001	85 (48)	3.1	3.7	0.23	.001
papG III	44 (15)	3.2	3.6	0.13	.03	17 (15)	3.0	3.4	0.12	.24	27 (15)	3.3	3.8	0.12	.04
sfa/focDE	71 (24)	3.1	3.7	0.20	.001	22 (19)	3.0	3.4	0.12	.21	49 (28)	3.2	3.9	0.22	<.001
sfaS	25 (9)	3.2	3.8	0.13	.02	11 (9)	3.1	3.3	0.06	.69	14 (8)	3.3	4.3	0.21	<.001
focG	34 (12)	3.2	3.6	0.08	.10	6 (5)	3.1	3.1	–0.02	1.0	28 (16)	3.4	3.7	0.08	.14
afa/draBC	21 (7)	3.3	3.5	0.05	.29	6 (5)	3.1	3.4	0.08	.61	15 (9)	3.4	3.6	0.05	.47
iha	102 (35)	3.1	3.5	0.15	.006	30 (26)	3.0	3.4	0.15	.14	72 (41)	3.3	3.6	0.11	.07
hra	86 (29)	3.2	3.5	0.13	.035	31 (27)	3.0	3.4	0.13	.13	55 (31)	3.3	3.6	0.11	.13
yfcV	245 (84)	2.7	3.4	0.20	<.001	97 (84)	2.5	3.2	0.21	.03	148 (84)	2.9	3.5	0.18	.03
Toxin															
hlyD	90 (31)	3.1	3.7	0.25	<.001	22 (19)	3.0	3.5	0.18	.07	68 (39)	3.2	3.8	0.27	<.001
hlyF	19 (7)	3.3	3.7	0.09	.08	5 (4)	3.1	4.0	0.13	.08	14 (8)	3.4	3.6	0.07	.47
cnf1	37 (13)	3.2	3.8	0.15	.009	1 (0.8)	3.1	3.1	–0.004	1.0	26 (15)	3.3	4.1	0.23	<.001
cdtB	18 (6)	3.3	3.6	0.06	.27	4 (3)	3.1	3.9	0.10	.40	14 (8)	3.4	3.5	0.01	.77
sat	94 (32)	3.1	3.6	0.18	.002	26 (22)	2.9	3.7	0.25	.009	68 (39)	3.3	3.6	0.10	.10
pic	31 (11)	3.2	3.8	0.14	.02	5 (4)	3.0	4.8	0.29	<.001	26 (15)	3.4	3.6	0.07	.38
vat	154 (53)	2.7	3.8	0.47	<.001	56 (48)	2.35	3.9	0.57	<.001	98 (56)	2.9	3.8	0.37	<.001
Siderophore															
iroN	90 (31)	3.1	3.8	0.26	<.001	30 (26)	2.9	3.8	0.31	.001	60 (34)	3.2	3.8	0.22	.002
fyuA	210 (72)	2.1	3.7	0.57	<.001	71 (61)	2.1	3.8	0.63	< .001	139 (79)	2.2	3.7	0.47	<.001
ireA	68 (23)	3.2	3.6	0.13	.03	23 (19)	2.9	3.8	0.24	.006	45 (26)	3.4	3.5	0.05	.63
iutA	126 (44)	3.1	3.6	0.15	.008	34 (29)	2.9	3.6	0.24	.01	92 (52)	3.3	3.5	0.05	.51
chuA	220 (75)	2.3	3.6	0.43	<.001	79 (68)	2.2	3.5	0.44	<.001	141 (80)	2.4	3.7	0.41	<.001

Table 2. Continued

	All Isolates (n = 292)					Fecal Isolates (n = 116)					Clinical Isolates (n = 176)				
		Mean Illness Severity Score					Mean Illness Severity Score					Mean Illness Severity Score			
Category, Trait[a]	No. (%)	Trait Absent	Trait Present	Correlation[b]	*P*[b]	No. (%)	Trait Absent	Trait Present	Correlation[b]	*P*[b]	No. (%)	Trait Absent	Trait Present	Correlation[b]	*P*[b]
Capsule															
kpsM II	211 (72)	2.3	3.7	0.49	<.001	78 (67)	.21	.36	0.50	<.001	133 (76)	2.4	3.7	0.46	<.001
K1	83 (28)	3.0	4.1	0.41	<.001	35 (30)	2.6	4.2	0.54	<.001	48 (27)	3.2	4.0	0.31	<.001
K2/K100	20 (7)	3.3	3.6	0.08	.19	1 (0.8)	3.1	2.0	–0.08	.36	19 (11)	3.4	3.7	0.11	.24
K5	32 (11)	3.2	3.7	0.11	.03	10 (9)	3.0	3.8	0.18	.049	22 (13)	3.4	3.6	0.05	.33
Miscellaneous															
cvaC	16 (5.4)	3.3	3.7	0.09	.10	5 (4)	3.0	4.4	0.18	<.001	11 (6)	3.4	3.4	0.07	.94
usp	169 (58)	2.6	3.8	0.48	<.001	64 (55)	2.2	3.8	0.59	<.001	105 (60)	2.8	3.8	0.40	<.001
traT	171 (59)	2.9	3.6	0.28	<.001	64 (55)	2.7	3.5	0.29	.001	107 (61)	3.0	3.6	0.23	.001
ibeA	47 (16)	3.1	4.1	0.28	<.001	18 (16)	3.0	3.9	0.24	.007	29 (16)	3.3	4.2	0.27	<.001
ompT	208 (71)	2.8	3.5	0.26	<.001	81 (70)	2.6	3.3	0.27	.003	127 (72)	2.9	3.6	0.24	.002
iss	21 (7)	3.2	3.8	0.11	.045	5 (4)	3.0	4.5	0.21	<.001	15 (90)	3.4	3.5	0.04	.65
H7 *fliC*	48 (16)	3.1	4.0	0.27	<.001	27 (23)	2.8	3.8	0.30	.001	21 (12)	3.3	4.3	0.29	<.001
malX	153 (52)	2.7	3.8	0.42	<.001	54 (47)	2.5	3.8	0.52	<.001	99 (56)	2.9	3.8	0.36	<.001
clbB/N	123 (42)	2.9	3.8	0.33/0.35	<.001	43 (37)	2.6	3.9	0.47/0.48	<.001	79 (45)	3.2	3.7	0.22/0.24	.001
Pathotype															
$ExPEC_{JJ}$	187 (64)	2.5	3.7	0.48	<.001	68 (59)	2.3	3.8	0.56	<.001	124 (70)	2.7	3.7	0.39	<.001
$UPEC_{HM}$	216 (74)	2.1	3.7	0.56	<.001	75 (65)	2.0	3.7	0.63	<.001	141 (80)	2.2	3.7	0.48	<.001
Source															
Clinical	176 (60)	3.1	3.4	0.12	.04	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Abbreviations: *E. coli*, *Escherichia coli*; NA, not applicable; STc, sequence type complex.

[a]Traits shown are those with a 5%–95% overall prevalence. Definitions are as follows: *papAH*, P fimbriae major subunit (results were similar for *papC*, *papEF*, and *papG*); *papG* allele II, variant P adhesin; *papG* allele III, variant P adhesin, *sfa/focDE*, S and F1C fimbriae; *sfaS*, S fimbriae; *focG*, F1C adhesin; *afa/draBC*, Dr-binding adhesins, *iha*, adhesin-siderophore; *yfcV*, chaperone-usher fimbriae; *hlyD*, hemolysin; *hlyF*, hemolysin F; *cnf1*, cytotoxic necrotizing factor 1; *cdtB*, cytolethal distending toxin; *sat*, secreted autotransporter toxin; *pic*, protein associated with intestinal colonization; *vat*, vacuolating autotransporter toxin; *iroN*, salmochelin receptor; *fyuA*, yersiniabactin receptor; *ireA*, siderophore receptor; *iutA*, aerobactin receptor; *chuA*, heme uptake; *kpsM* II, group 2 capsules; K1, variant group 2 capsule; K2/K100, group 2 capsule variants; K5, group 2 capsule variant; *cvaC*, colicin (microcin) V; *usp*, uropathogenic-specific protein; *traT*, serum resistance-associated surface exclusion protein; *ibeA*, invasion of brain endothelium; *ompT*, outer membrane protease; *iss*, increased serum survival; H7 *fliC*, variant flagellin; *malX*, pathogenicity island marker; *clbB/N*, colibactin synthesis; $ExPEC_{JJ}$, extraintestinal pathogenic *E. coli* (molecular definition per James Johnson: ≥2 of *papA* and/or *papC*, *sfa/foc*, *afa/dra*, *kpsM* II, and *iutA*); $UPEC_{HM}$, uropathogenic *E. coli* (molecular definition per Harry Mobley: ≥2 of *chuA*, *fyuA*, *vat*, and *yfcV*).

[b]Spearman correlation coefficient (ρ).

[c]By a 2-tailed *t* test. Values <.05 are considered statistically significant.

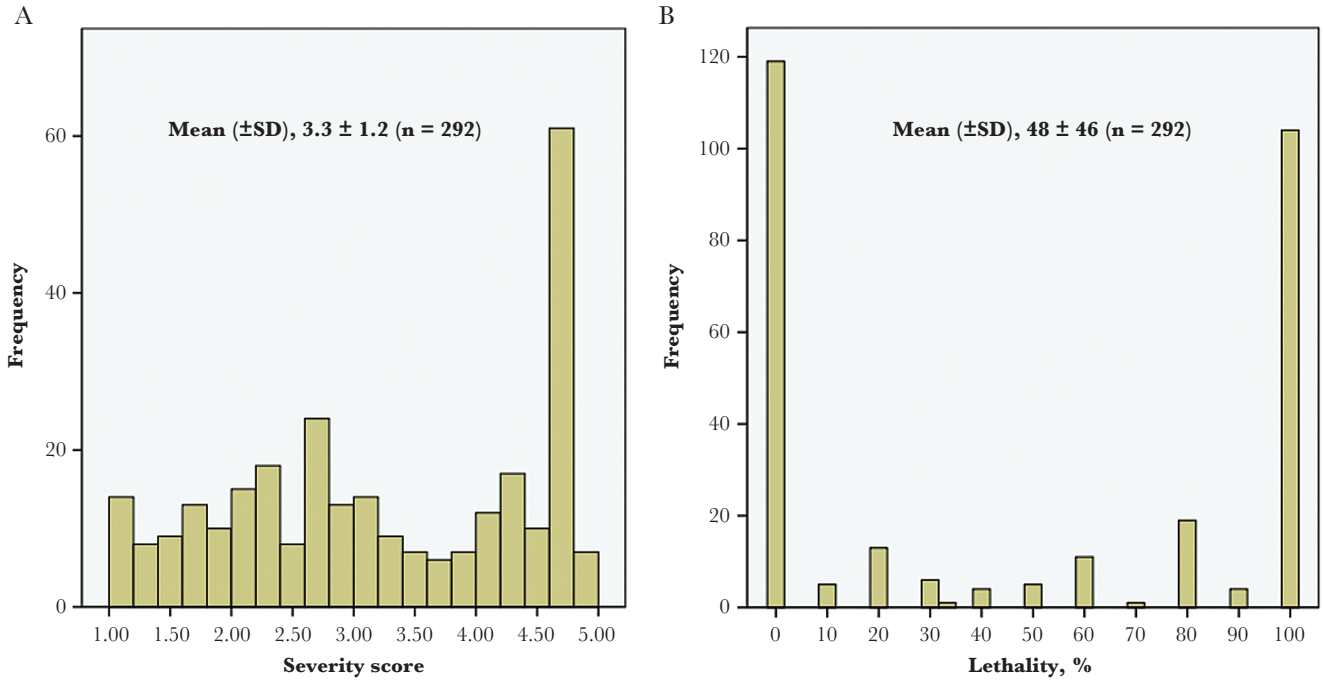

Figure 1. Distribution of mean illness severity (*A*) and lethality percentage (*B*) in the murine sepsis model for 292 clinical (ie, urine and blood) and fecal *Escherichia coli* isolates.

Univariable Comparisons of Bacterial Characteristics and Source to Virulence

Of the studied bacterial characteristics, 46 qualified for statistical analysis, based on a prevalence of 5%–95% (Table 2). These included 4 phylogenetic groups, 5 STc, 35 accessory traits (ie, 10 adhesin genes, 7 toxin genes, 5 siderophore systems, 4 capsule markers, and 9 miscellaneous traits), and 2 composite variables (ie, $ExPEC_{JJ}$ and $UPEC_{HM}$). These bacterial characteristics, plus clinical (vs fecal) source, were compared to illness severity (Table 2).

Overall, 38 analyzed bacterial traits (83%), including multiple individual traits per category plus the composite variables, $ExPEC_{JJ}$ and $UPEC_{HM}$, were correlated significantly with illness severity in ≥1 (usually all 3) population partitions (Table 2). Most correlations were positive and highly statistically significant and, when significant among both fecal and clinical isolates, pointed in the same direction in each group. Of the 8 exceptional traits that were uncorrelated with illness severity, 6 exhibited an overall prevalence of <10%.

Clinical source also was correlated significantly with illness severity, albeit weakly ($\rho = 0.12$; $P = .04$; Table 2). Mean severity scores (±SD) overlapped considerably between the fecal and clinical isolates (3.1 ± 1.3 and 3.4 ± 1.2, respectively). Other population partitions by source (ie, invasive vs noninvasive and blood vs nonblood) yielded no significant correlations with illness severity (data not shown).

Derived Sets of Predictors of Illness Severity

The multistage process that was used to select variables for multivariable modeling identified, within the total population, 6 bacterial characteristics as primary variables ($\rho < .70$; Table 3). Neither $UPEC_{HM}$ nor $ExPEC_{JJ}$ qualified, since both were highly correlated with (stronger predictor) *fyuA*.

Among the fecal isolates, this process identified 9 primary variables, including 5 of 6 from the total population analysis (Table 3). *usp*, a primary variable in the total population analysis, was excluded here because it was correlated with (ie, was a stronger predictor of) *fyuA*. When assessed together with the individual bacterial traits, $UPEC_{HM}$ and $ExPEC_{JJ}$ now qualified as primary variables, along with *vat*, in place of 3 of the initial primary variables (ie, *fyuA*, *clbN*, and *kpsM* II), which were dominated by the correlated variables (ie, stronger predictors) $UPEC_{HM}$, *vat*, and $ExPEC_{JJ}$, respectively (Table 3).

Among the clinical isolates, this process identified 5 primary variables, including 3 that had been identified previously as primary variables (ie, *fyuA*, *kpsM* II, and K1) and one that was interchangeable with such variables (ie, *papEF* vs *papAH*); the only newly identified variable was group B2. As in the total population analysis, neither $UPEC_{HM}$ nor $ExPEC_{JJ}$ qualified because they were highly correlated with (stronger predictor) *fyuA*.

Multiple Regression Models

To identify independent predictors of illness severity, these derived sets of variables (Table 3) were used as candidate predictors in multiple regression models (Table 4). Forward and backward stepwise entry consistently arrived at the same final result (data not shown); hence, Table 4 shows the results of forward stepwise entry only.

Table 3. Intercorrelations Among Bacterial Variables Significantly Associated With Illness Severity

		Variable (ρ for Correlation With Illness Severity)	
Population	$UPEC_{HM}$/$ExPEC_{JJ}$ Considered?[a]	Primary Variable[b]	Correlated Variables,[b] Dominated by Primary Variable
Total (n = 292)	No or yes	*fyuA* (0.57)	Without $UPEC_{HM}$/$ExPEC_{JJ}$, none; with them, $UPEC_{HM}$ (0.56), $ExPEC_{JJ}$ (0.48)
		kpsM II (0.49)	*chuA* (0.43)
		usp (0.48)	*vat* (0.46), group B2 (0.46), *malX* (0.43), *clbB* (0.33), *clbN* (0.35)
		K1 (0.41)	None
		papAH (0.37)	*papEF* (0.37), *papC* (0.37), *papG* (0.33)
		group B1 (–0.31)	None
Fecal (n = 116)	No	*fyuA* (0.63)	*usp* (0.59), *vat* (0.57), group B2 (0.52)
		K1 (0.54)	None
		malX (0.52)	None
		kpsM II (0.50)	*chuA* (0.45)
		clbN (0.48)	*clbB* (0.47)
		group B1 (–0.40)	None
		STc95 (0.36)	None
		papAH (0.35)	*papEF* (0.34), *papC* (0.33)
		iroN (0.31)	None
	Yes	$UPEC_{HM}$ (0.63)	*fyuA* (0.63), *usp* (0.59), group B2 (0.52)
		vat (0.57)	*clbN* (0.48), *clbB* (0.48)
		$ExPEC_{JJ}$ (0.56)	*kpsM* II (0.50)
		malX (0.52)	None
		group B1 (–0.40)	None
		STc95 (0.36)	None
		papAH (0.35)	*papEF* (0.37), *papC* (0.33)
		iroN (0.31)	None
		K1 (0.31)	None
Clinical (n = 176)	No or yes	*fyuA* (0.49)	Without $UPEC_{HM}$/$ExPEC_{JJ}$: none; with them: $UPEC_{HM}$ (0.48), $ExPEC_{JJ}$ (0.39)
		kpsM II (0.46)	*chuA* (0.41)
		group B2 (0.40)[c]	*usp* (0.40),[c] *vat* (0.37), *malX* (0.36)
		papEF (0.36)	*papC* (0.36), *papAH* (0.35), *papG* (0.35)
		K1 (0.31)	None

ρ values of ≥ 0.30 were considered indicative of a statistical association with illness severity.

Abbreviations: $ExPEC_{JJ}$, extraintestinal pathogenic *E. coli*, as defined using James Johnson's molecular definition; $UPEC_{HM}$, uropathogenic *E. coli*, as defined using Harry Mobley's molecular definition.

[a]The analysis was done with and without considering $UPEC_{HM}$ and $ExPEC_{JJ}$ as candidate predictor variables.

[b]Definitions are as follows: *fyuA*, yersiniabactin receptor; *kpsM* II, group 2 capsule synthesis; *usp*, uropathogenic specific protein; K1, group 2 capsule variant; *papAH, papC, papEF, papG*, P fimbriae structural subunit, assembly, minor tip pilins, and tip adhesin; groups B1 and B2, phylogenetic groups; *malX*, pathogenicity island marker; *clbN* and *clbB*, colibactin synthesis; STc95, sequence type complex 95; *iroN*, salmochelin receptor; *vat*, vacuolating toxin; *malX*, pathogenicity island marker; *chuA*, heme uptake.

[c]Among the clinical isolates, group B2 and *usp* yielded identical values for rho (0.398), but B2 dominated two alternate correlated variables (*vat* and *malX*), whereas *usp* dominated only one (*malX*). Accordingly, for parsimony, B2 was selected over *usp* for inclusion in the model.

Within the total population, in univariable models $UPEC_{HM}$ yielded an r^2 of 0.31 and $ExPEC_{JJ}$ yielded an r^2 of 0.23, whereas the multivariable forced entry model yielded an r^2 of 0.39 and identified as significant 2 of 6 candidate predictor variables: *fyuA* (β = 0.35; $P < .001$) and K1 (β = 0.21; $P < .001$; Table 4). By contrast, the stepwise models identified also *kpsM* II (β = 0.15; $P = .02$) and achieved an r^2 of 0.39 with only 3 predictor variables. Bootstrapping results were confirmatory.

Among the fecal isolates, in univariable models $UPEC_{HM}$ yielded an r^2 of 0.39 and $ExPEC_{JJ}$ yielded an r^2 of 0.31, whereas without $UPEC_{HM}$ and $ExPEC_{JJ}$ the multivariable forced entry model yielded an r^2 of 0.46 and identified as significant 2 of 9 candidate predictor variables: K1 (β = 0.33; $P < .001$) and *fyuA* (β = 0.30; $P = .01$; Table 4). The corresponding stepwise models likewise achieved an r^2 of 0.46 and identified the same 2 variables as significant predictors, in reverse order of potency. By contrast, with $UPEC_{HM}$ and $ExPEC_{JJ}$ included, the forced entry model achieved a slightly higher r^2 (r^2 = 0.48) and identified as significant predictors K1 (β = 0.33; $P = .003$) and $UPEC_{HM}$ (β = 0.28; $P = .01$), whereas the corresponding stepwise models yielded an r^2 of 0.49 and identified $ExPEC_{JJ}$ (β = 0.19, $P = .05$) as a significant predictor, as well. Bootstrapping findings were confirmatory except with regard to the significance of $ExPEC_{JJ}$ ($P = .08$).

Finally, among the clinical isolates, univariable models $UPEC_{HM}$ yielded an r^2 of 0.23 and $ExPEC_{JJ}$ yielded an r^2 of 0.15, whereas the multivariable forced entry model yielded an r^2 of 0.32 and identified as significant 2 of 5 primary predictor

Table 4. Multivariable Models to Predict Illness Severity in Mice Challenged with *Escherichia coli* Isolates

E. coli Set	$UPEC_{HM}$/$ExPEC_{JJ}$ Considered?[a]	Method[b]	Model No.	Adjusted r^2	Variable[c]	β	*P*[d]
Total (n = 292)	No or yes[e]	Forced	NA[f]	.39	*fyuA*	0.35	<.001
					K1	0.21	<.001
					kpsM II	0.11	.10
					papAH	0.06	.26
					usp	0.06	.40
					Group B1	0.003	.96
		Stepwise	1	.33	*fyuA*	0.58	<.001
			2	.38	*fyuA*	0.49	<.001
					K1	0.24	<.001
			3	.39	*fyuA*	0.40	<.001
					K1	0.21	<.001
					kpsM II	0.15	.02
Fecal (n = 116)	No	Forced	NA[f]	.46	K1	0.33	.002
					fyuA	0.30	.01
					malX	0.13	.20
					papAH	–0.11	.26
					clbN	0.10	.36
					iroN	0.06	.46
					kpsM II	0.06	.61
					Group B1	–0.02	.79
					STc95	–0.01	.90
		Stepwise	1	.39	*fyuA*	0.63	<.001
			2	.46	*fyuA*	0.47	<.001
					K1	0.32	<.001
	Yes	Forced	NA[f]	.48	K1	0.33	.003
					$UPEC_{HM}$	0.28	.01
					$ExPEC_{JJ}$	0.22	.08
					papAH	–0.16	.14
					malX	0.10	.32
					iroN	0.07	.42
					vat	0.03	.81
					Group B1	–0.006	.95
					STc95	–0.002	.98
		Stepwise	1	.39	$UPEC_{HM}$	0.63	<.001
			2	.47	$UPEC_{HM}$	0.47	<.001
					K1	0.34	<.001
			3	.49	$UPEC_{HM}$	0.35	.001
					K1	0.32	<.001
					$ExPEC_{JJ}$	0.19	.05[d]
Clinical (n = 176)	No or yes [a]	Forced	NA[f]	.32	*fyuA*	0.29	.002
					K1	0.17	.01
					kpsM II	0.17	.06
					papEF	0.13	.10
					Group B2	0.01	.89
		Step	1	.26	*fyuA*	0.51	<.001
			2	.30	*fyuA*	0.35	<.001
					kpsM II	0.26	.002
			3	.31	*fyuA*	0.34	<.001
					kpsM II	0.21	.01
					K1	0.16	.02

Abbreviations: $ExPEC_{JJ}$, extraintestinal pathogenic *E. coli*, as defined using James Johnson's molecular definition; NA, not applicable; $UPEC_{HM}$, uropathogenic *E. coli*, as defined using Harry Mobley's molecular definition.

[a]The analysis was done with and without considering $UPEC_{HM}$ and $ExPEC_{JJ}$ as candidate predictor variables.

[b]Forced variable entry or conditional stepwise variable entry. Forward and backward stepwise entry gave identical final models; for brevity, only the stepwise forward entry models are shown.

[c]Definitions are specified in Tables 2 and 3.

[d]Values <.05 are considered statistically significant. Bootstrapping findings supported all significant *P* values except that for ExPEC (with fecal isolates, $UPEC_{HM}$/$ExPEC_{JJ}$ were considered in a stepwise model).

[e]Results were the same regardless of whether $UPEC_{HM}$ and $ExPEC_{JJ}$ were considered, since they did not qualify as predictor variables.

[f]For forced entry, there was only 1 model per data set and variable list.

variables: *fyuA* (β = 0.29; *P* = .002) and K1 (β = 0.17; *P* = .01; Table 4). Additionally, *kpsM* II approached statistical significance (β = 0.17; *P* = .06). The corresponding stepwise models achieved an r^2 of 0.31 and identified all 3 variables as significant. Bootstrapping findings were confirmatory.

For exploratory purposes, clinical source was added to the (total population) multivariable models as a candidate predictor, despite yielding a ρ of <0.30 with illness severity (Table 2). In the forced entry model, it was nonsignificant and did not increase r^2, and in the stepwise models it was excluded (data not shown).

DISCUSSION

Here we assessed ecological source and diverse bacterial traits as predictors of virulence in a murine sepsis model for 292 clinical and fecal *E. coli* isolates. Our findings support 4 main conclusions. First, multiple bacterial characteristics, including specific phylogenetic groups, clonal groups (ie, STc), and accessory traits (ie, virulence genes), significantly predicted experimental virulence. Second, trait combinations were more predictive than any single trait, explaining up to 49% of the total virulence. Third, ecological (ie, clinical vs fecal) source, when considered in isolation, also significantly predicted experimental virulence. Fourth, bacterial traits were much more potent predictors than was ecological source, and similar traits were predictive among clinical and fecal isolates.

These conclusions, which support and extend previously proposed concepts [11, 25], have important implications regarding the relationship between intrinsic virulence, bacterial traits, and ecological source. A common approach for classifying *E. coli* isolates on the basis of their presumed extraintestinal virulence potential relies on source: clinical isolates presumptively are virulent pathogens (eg, UPEC or ExPEC), whereas fecal isolates presumptively are low-virulence commensal organisms [3, 5]. By contrast, we found a broad range of experimental virulence among clinical and fecal isolates alike, demonstrating that an isolate's clinical versus fecal origin does not indicate reliably its intrinsic virulence potential. Although on average the clinical isolates (which represented predominantly pyelonephritis and bacteremia) tended to be more virulent than the fecal isolates, these populations overlapped considerably.

We selected our study population by deliberately stratifying isolates according molecular $ExPEC_{JJ}$ status to provide sufficient numbers within key subgroups to overcome the statistical power limitations that occur with natural or ad hoc isolate collections. Such biased selection might be considered to preclude a valid assessment of virulence in relation to ecologic source by distorting the authentic relationship between source and $ExPEC_{JJ}$ status. However, this concern actually presupposes the validity of the study's conclusion that genetic content is a more important determinant of experimental virulence than is ecological source. Here, ecological source was so weak a predictor of illness severity that it did not qualify for inclusion in the multivariable models and, when included anyway, proved noncontributory. Additional evidence of virulence commonality across source groups was the finding that similar trait combinations, including the established composite variables, $ExPEC_{JJ}$ and $UPEC_{HM}$, predicted experimental virulence among fecal and clinical isolates alike.

Collectively, these findings suggest that the observed overall virulence differences between clinical and fecal isolates can be explained better by these populations' different admixtures of virulent versus nonvirulent strains, rather than by categorical strain-level virulence differences by source. The most strongly predictive individual traits, both overall and among fecal and clinical isolates, were *fyuA* (yersiniabactin system) and *kpsM* II and K1 (group 2 capsule genes), although to what extent they contributed directly to or were simply markers of virulence is unclear.

Nonetheless, even the best-performing trait-based prediction models yielded an r^2 of <50%. To the extent that bacterial characteristics determine infection outcomes, this implies the possible importance of traits other than those we tested, perhaps including known virulence genes [26] or as-yet-unknown traits that await discovery, such as by genome [27, 28] or transcriptome [29] analysis. Alternatively, expression levels, trait combinations, multiply determined phenotypes, and/or regulatory pathways may be important, as may be other typing methods or advanced analytical approaches [30].

Host factors also must be considered. In humans, host compromise allows low-virulence *E. coli* strains to cause severe disease [6, 31]. Likewise, different inbred mouse strains with specific immune polymorphisms exhibit different responses to bacterial challenge [32, 33]. Host-to-host variation may be relevant here since we used outbred mice, for which the host response to bacterial challenge may vary by mouse. To avoid confounding by such host variation, we assigned mice randomly to each test strain. Nonetheless, stochastic effects could have led to some strains being administered to groups of mice that were disproportionately more or less resistant to infection, biasing the results for those strains.

The present findings' relevance to *E. coli* infections in humans is unknown. Mice are not humans, and the model is highly artificial. Moreover, given known differences in anatomic site–specific host defenses, nutrient resources, and bacterial receptors, different animal models might yield different results. Still, given the appreciable commonality across anatomic sites with respect to the challenges pathogens face in persisting and causing disease, certain bacterial characteristics likely are important, regardless of the specific model or host species. Notably, our nonurinary tract model identified as significant virulence predictors certain traits, such as *pap* and *usp,* that were interpreted initially—and even named ("pyelonephritis-associated pili" and "uropathogenic-specific protein," respectively)—as

urovirulence factors [34, 35] and that have proven urinary tract–specific mechanisms [36, 37]. Likewise, $UPEC_{HM}$, a composite variable derived by molecular epidemiological comparisons of human urine and fecal isolates, here was predictive of systemic virulence, especially so among fecal isolates, in a model that bypasses completely the urinary tract. This provides further evidence of the tenuous and somewhat arbitrary nature of the distinction between UPEC versus ExPEC and between urovirulence factors versus generic extraintestinal virulence factors.

Our use of a broader pool of candidate predictor variables than studied previously identified novel trait combinations that outperformed the established molecular definitions ExPEC [18] and UPEC [3], suggesting the possibility of devising improved molecular definitions. However, for this it would be desirable to test additional isolates, including those from humans, animals [38], food [39], and the environment [40], in multiple infection models, and to supplement such experimental data with epidemiological data that, ideally, would consider host compromise status and clinical presentation.

Study limitations include the reliance on a murine sepsis model, analysis only of selected bacterial traits (and for presence or absence only), inattention to host factors, and small numbers in some subgroups. Study strengths include the comparatively large study population and broad range of traits analyzed, the multistage analytical approach, and the inclusion of the composite variables, $ExPEC_{JJ}$ and $UPEC_{HM}$.

In summary, the experimental virulence in mice of human clinical and fecal *E. coli* isolates corresponded more closely with virulence gene content and phylogenetic background than with source, and even the most predictive combinations of bacterial variables explained less than half of the observed virulence variation. This both confirms the primacy of bacterial traits over ecological origin in predicting the extraintestinal virulence in *E. coli* and identifies a need to investigate further the host and bacterial determinants of extraintestinal virulence.

Notes

Acknowledgments. We thank the study participants and the contributing clinical microbiology laboratories.

Disclaimer. The opinions expressed are strictly those of the authors and not those of their respective institutions, the Department of Veterans Affairs, or the National Institutes of Health.

Financial support. This work was supported in by the Office of Research and Development, Department of Veterans Affairs (grants 1 I01 CX000920-01 and 2I01CX000920-04 to J. R. J.) and the National Institutes of Health (grant R21AI117654 to L. B. P.).

Potential conflicts of interest. J. R. J. has received research grants from Achaogen, Allergan, Melinta, Merck, Syntiron, and Tetraphase; collaborates with IDGenomics; is a consultant for Crucell/Janssen and Syntiron; and has patent applications for tests to detect specific *E. coli* strains. L. B. P. has a patent application for tests to detect specific *E. coli* strains and a grant from Merck. All other authors report no potential 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.

References

1. Russo TA, Johnson JR. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: an overlooked epidemic. Microbes Infect **2003**; 5:449–56.
2. Russo TA, Johnson JR. A proposal for an inclusive designation for extraintestinal pathogenic *Escherichia coli*: ExPEC. J Infect Dis **2000**; 181:1753–4.
3. Spurbeck RR, Dinh PC Jr, Walk ST, et al. *Escherichia coli* isolates that carry *vat*, *fyuA*, *chuA*, and *yfcV* efficiently colonize the urinary tract. Infect Immun **2012**; 80:4115–22.
4. Moreno E, Andreu A, Pigrau C, Kuskowski MA, Johnson JR, Prats G. 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.
5. Sannes MR, Kuskowski MA, Owens K, Gajewski A, Johnson JR. Virulence factor profiles and phylogenetic background of *Escherichia coli* isolates from veterans with bacteremia versus uninfected control patients. J Infect Dis **2004**; 190:2121–8.
6. Johnson JR, Porter S, Johnston B, et al. Host characteristics and bacterial traits predict experimental virulence for *Escherichia coli* bloodstream isolates from patients with urosepsis. Open Forum Infect Dis **2015**; 2:ofv083.
7. de Man P, Cläeson I, Johanson IM, Jodal U, Svanborg Edén C. Bacterial attachment as a predictor of renal abnormalities in boys with urinary tract infection. J Pediatr **1989**; 115:915–22.
8. Yamamoto S, Tsukamoto T, Terai A, Kurazono H, Takeda Y, Yoshida O. Genetic evidence supporting the fecal-perineal-urethral hypothesis in cystitis caused by *Escherichia coli*. J Urol **1997**; 157:1127–9.
9. 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.
10. Wold AE, Caugant DA, Lidin-Janson G, de Man P, Svanborg C. Resident colonic *Escherichia coli* strains frequently display uropathogenic characteristics. J Infect Dis **1992**; 165:46–52.
11. Johnson JR, Clermont O, Menard M, Kuskowski MA, Picard B, Denamur E. Experimental mouse lethality of *Escherichia coli* isolates, in relation to accessory traits, phylogenetic group, and ecological source. J Infect Dis **2006**; 194:1141–50.

12. Diard M, Baeriswyl S, Clermont O, et al. *Caenorhabditis elegans* as a simple model to study phenotypic and genetic virulence determinants of extraintestinal pathogenic *Escherichia coli*. Microbes Infect **2007**; 9:214–23.
13. Lavigne JP, Vergunst AC, Goret L, et al. Virulence potential and genomic mapping of the worldwide clone *Escherichia coli* ST131. PLoS One **2012**; 7:e34294.
14. Picard B, Garcia JS, Gouriou S, et al. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. Infect Immun **1999**; 67:546–53.
15. Sannes MR, Kuskowski MA, Johnson JR. Antimicrobial resistance among *Escherichia coli* isolates from women with cystitis and pyelonephritis, healthy human volunteers, and domestic dog feces. J Am Vet Med Assoc **2004**; 225:368–73.
16. Sannes MR, Kuskowski MA, Johnson JR. Geographical distribution of antimicrobial resistance among *Escherichia coli* causing acute uncomplicated pyelonephritis in the United States. FEMS Immunol Med Microbiol **2004**; 42:213–8.
17. Johnson JR, Orskov I, Orskov F, et al. O, K, and H antigens predict virulence factors, carboxylesterase B pattern, antimicrobial resistance, and host compromise among *Escherichia coli* strains causing urosepsis. J Infect Dis **1994**; 169:119–26.
18. 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.
19. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **2009**; 25:1754–60.
20. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res **2010**; 20:1297–303.
21. Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res **2015**; 43:e15.
22. Guindon S, Lethiec F, Duroux P, Gascuel O. PHYML Online–a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res **2005**; 33:W557–9.
23. 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.
24. Johnson JR, Porter SB, Zhanel G, Kuskowski MA, Denamur E. Virulence of *Escherichia coli* clinical isolates in a murine sepsis model in relation to sequence type ST131 status, fluoroquinolone resistance, and virulence genotype. Infect Immun **2012**; 80:1554–62.
25. Johnson JR, Kuskowski M, Denamur E, Elion J, Picard B. Clonal origin, virulence factors, and virulence (letter and reply). Infect Immun **2000**; 68:424–5.
26. Spurbeck RR, Stapleton AE, Johnson JR, Walk ST, Hooton TM, Mobley HL. Fimbrial profiles predict virulence of uropathogenic *Escherichia coli* strains: contribution of *ygi* and *yad* fimbriae. Infect Immun **2011**; 79:4753–63.
27. Salipante SJ, Roach DJ, Kitzman JO, et al. Large-scale genomic sequencing of extraintestinal pathogenic *Escherichia coli* strains. Genome Res **2015**; 25:119–28.
28. McNally A, Oren Y, Kelly D, et al. Combined analysis of variation in core, accessory and regulatory genome regions provides a super-resolution view into the evolution of bacterial populations. PLoS Genet **2016**; 12:e1006280.
29. Kisiela DI, Radey M, Paul S, et al. Inactivation of transcriptional regulators during within-household evolution of *Escherichia coli*. J Bacteriol **2017**:(in press).
30. Bakhtiarizadeh MR, Moradi-Shahrbabak M, Ebrahimi M, Ebrahimie E. Neural network and SVM classifiers accurately predict lipid binding proteins, irrespective of sequence homology. J Theor Biol **2014**; 356:213–22.
31. Lomberg H, Hanson LA, Jacobsson B, Jodal U, Leffler H, Edén CS. Correlation of P blood group, vesicoureteral reflux, and bacterial attachment in patients with recurrent pyelonephritis. N Engl J Med **1983**; 308:1189–92.
32. Godaly G, Ambite I, Svanborg C. Innate immunity and genetic determinants of urinary tract infection susceptibility. Curr Opin Infect Dis **2015**; 28:88–96.
33. Hagberg L, Hull R, Hull S, McGhee JR, Michalek SM, Svanborg Edén C. Difference in susceptibility to Gram-negative urinary tract infection between C3H/HeJ and C3H/HeN mice. Infect Immun **1984**; 46:839–44.
34. Hull RA, Hull SI, Falkow S. Frequency of gene sequences necessary for pyelonephritis-associated pili expression among isolates of Enterobacteriaceae from human extraintestinal infections. Infect Immun **1984**; 43:1064–7.
35. Kurazono H, Yamamoto S, Nakano M, et al. Characterization of a putative virulence island in the chromosome of uropathogenic *Escherichia coli* possessing a gene encoding a uropathogenic-specific protein. Microb Pathog **2000**; 28:183–9.
36. Korhonen TK, Virkola R, Holthöfer H. Localization of binding sites for purified *Escherichia coli* P fimbriae in the human kidney. Infect Immun **1986**; 54:328–32.
37. Yamamoto S, Nakano M, Terai A, et al. The presence of the virulence island containing the usp gene in uropathogenic *Escherichia coli* is associated with urinary tract infection in an experimental mouse model. J Urol **2001**; 165:1347–51.
38. Johnson TJ, Wannemuehler Y, Johnson SJ, et al. Comparison of extraintestinal pathogenic *Escherichia coli* strains from human and avian sources reveals a mixed subset representing potential zoonotic pathogens. Appl Environ Microbiol **2008**; 74:7043–50.

39. Johnson JR, Porter SB, Johnston B, et al. Extraintestinal pathogenic and antimicrobial resistant *Escherichia coli*, including sequence type 131 (ST131), from retail chicken breasts: United States, 2013. Appl Environ Microbiol **2017**; 83. doi: 10.1128/AEM.02956-16.

40. Mohamed M, Owens K, Gajewski A, et al. Extraintestinal pathogenic and antimicrobial-resistant *Escherichia coli* Contamination of 56 public restrooms in the greater Minneapolis-St. Paul metropolitan area. Appl Environ Microbiol **2015**; 81:4498–506.