MAJOR ARTICLE

Experimental Mouse Lethality of *Escherichia coli* Isolates, in Relation to Accessory Traits, Phylogenetic Group, and Ecological Source

James R. Johnson,[1,3] Olivier Clermont,[5] Megan Menard,[1,3] Michael A. Kuskowski,[2,4] Bertrand Picard,[5] and Erick Denamur[5]

[1]Mucosal and Vaccine Research Center and [2]Geriatric Research, Education, and Clinical Center, VA Medical Center, and Departments of [3]Medicine and [4]Psychiatry, University of Minnesota, Minneapolis; [5]INSERM U722 and Université Paris 7, Paris, France

Background. Whether accessory traits, phylogenetic background, or ecological source best predicts extraintestinal virulence within *Escherichia coli* is undefined.

Methods. A total of 90 *E. coli* strains (18 fecal isolates and 72 extraintestinal-infection isolates) were characterized for 55 accessory traits and phylogenetic group (A, B1, B2, or D). Bacterial traits and ecological source were compared with experimental mouse lethality.

Results. Of the 90 strains, 41% were "killers" (i.e., killed ⩾90% of mice). By univariate analysis, multiple group B2–associated traits (including *malX* [pathogenicity-island marker], *pap* [P fimbriae] elements, *usp* [uropathogenic-specific protein], and *fyuA* [yersiniabactin system]) were most closely associated with killer status, followed by group B2 (or non–group A) status and then by nonfecal origin. Stepwise multivariate analysis identified *pap*, *malX*, *usp*, *fyuA*, and B2 (all of which were positive predictors) and *ireA* (which was a negative predictor) as significant predictors of killer status. Killer strains segregated significantly from nonkiller strains, according to accessory-trait profiles. Factorial analysis of correspondence placed group B2 among the traits most closely associated with killer status, but not as the closest.

Conclusions. Specific group B2–associated accessory traits are more potent predictors of experimental virulence among *E. coli* isolates than is either phylogenetic background or ecological source. Molecular typing can estimate an *E. coli* isolate's extraintestinal virulence potential, regardless of source.

Extraintestinal infections—including urinary tract infections, meningitis, and sepsis—caused by *Escherichia coli* are a major cause of morbidity, mortality, and increased health-care costs [1, 2]. The causative extraintestinal pathogenic *E. coli* (ExPEC) strains typically belong to phylogenetic group B2 and express diverse accessory traits—such as polysaccharide coatings, toxins, adhesins, siderophores, and invasins—that distinguish them from commensal *E. coli* and that may contribute directly to virulence [3, 4]. Such traits, or the underlying genes, serve as useful epidemiological markers in studies assessing the presumed virulence of different *E. coli* populations or individual isolates [5] and are candidate targets for preventive or therapeutic interventions [6]. However, it is unknown which accessory traits are most closely linked with the ability to cause disease—and whether accessory traits, phylogenetic background, or ecological source is most predictive of virulence potential [7, 8].

Assessment of specific bacterial traits as markers for or contributors to extraintestinal virulence usually is accomplished either by observational comparisons of carefully selected populations or by experimental analysis of individual *E. coli* strains [5, 9]. The epidemiological approach assesses "real world" phenomena and permits attention to multiple traits simultaneously. However, it is vulnerable to confounding from unmeasured differences between host groups and from

Received 2 December 2005; accepted 19 May 2006; electronically published 15 September 2006.

Potential conflicts of interest: J.R.J. has received grant support and/or honoraria from Bayer, Wyeth-Ayerst, Ortho-McNeil, Merck, Procter and Gamble, and Rochester Medical Group.

Financial support: Office of Research and Development, Medical Research Service, Department of Veterans Affairs (J.R.J.); Fondation pour la Recherche Médicale (E.D.).

Reprints or correspondence: Dr. James R. Johnson, Infectious Diseases (111F), Minneapolis VA Medical Ctr., 1 Veterans Dr., Minneapolis, MN 55417 (johns007@umn.edu).

The Journal of Infectious Diseases 2006; 194:1141–50

0022-1899/2006/19408-0016$15.00

misclassification (e.g., the presence of high-virulence *E. coli* within fecal populations [10] or of low-virulence *E. coli* among clinical isolates from compromised hosts [11]). In contrast, the experimental approach typically involves analysis of a single bacterial trait in a controlled animal-model system. This provides considerable confidence regarding the internal validity of the findings [9]; however, it leaves uncertain their applicability to naturally occurring human infections, including those involving heterologous *E. coli* strains, and offers little opportunity for assessment of multiple bacterial traits either simultaneously or in comparison with one another.

Accordingly, we used a third approach—namely, experimental-virulence screening of a large panel of extensively characterized *E. coli* isolates—to assess, simultaneously, multiple accessory traits, phylogenetic group, and ecological source, as predictors of extraintestinal-virulence potential [12, 13]. Specifically, we used polymerase chain reaction (PCR)–based assays to characterize 90 clinical and fecal *E. coli* strains for 55 accessory traits and major *E. coli* phylogenetic group (A, B1, B2, or D) [14]. We then compared molecular-typing results and the isolate's ecological source with results from a standardized mouse-lethality assay [12], using a variety of analytical approaches to identify the most potent predictors of experimental virulence [13].

MATERIALS AND METHODS

Experimentation guidelines of the authors' institutions were followed in the conduct of clinical research.

Strains. The 90 *E. coli* strains studied comprised all 79 available isolates (14 fecal and 65 extraintestinal infection isolates) among the 82 isolates from a previous similar study [12], plus 11 new strains. The 11 new strains included 4 archetypal ExPEC isolates (J96, 536, CFT073, and RS218) [15–18], 3 other extraintestinal isolates (EC7372, F11, and F63), 3 group B2 fecal isolates (Ben13f, colF6c, and ED1a), and *E. coli* K-12 derivative MG1655 [19]. MG1655 was analyzed as a fecal isolate, resulting in a total of 18 fecal isolates. The 72 extraintestinal isolates, hereafter termed "infection isolates," were from urine (57 isolates), blood (6 isolates), and other sources (9 isolates) and were from patients with extraintestinal *E. coli* infection. The 18 fecal isolates were from uninfected hosts. The "ecological source" variable referred to an isolate's fecal versus extraintestinal-infection origin.

Phylotyping and accessory traits. *E. coli* phylogenetic group (A, B1, B2, or D) and 55 putative virulence-associated accessory traits, including 3 alleles of *papG* (P fimbriae adhesin) and 13 alleles of *papA* (P fimbriae structural subunit), were identified by established multiplex PCR assays [14, 20–23]. All strains were tested at least in duplicate, with appropriate positive and negative controls, by use of boiled lysates as template DNA.

Mouse-lethality assay. A mouse subcutaneous–infection model was used to assess the lethality of the *E. coli* strains [12]. For each strain, 10–20 outbred female Swiss mice (6–8 weeks old and weighing 25–30 gm) were challenged subcutaneously, in the abdomen, with a standardized bacterial inoculum (10^9 cfu/mL log-phase bacteria in 0.2 mL Ringer solution). Lethality was assessed for 7 days postchallenge. (In this model, subcutaneous injection of 10^8 cfu allows the recovery of bacteria, in concentrations $\geqslant 10^8$ cfu/mL, from the liver, spleen, and kidneys of killed mice, whereas the organs of nonkilled mice are sterile [12].) Each experimental series included a positive control (urosepsis strain CFT073; all mice were killed by 7 days postchallenge) and 2 negative controls (one inoculated with Ringer solution, the other inoculated with strain MG1655; no mice were killed by 7 days postchallenge).

For previously tested strains, existing lethality data were used [12]. For the new strains, new testing was done to determine the proportion of mice dying by 7 days postchallenge. In this model, lethality is a clear-cut parameter, with >80% of strains that kill any mice killing ⩾90% of mice [12]. Because (1) results were similar whether lethality was analyzed either dichotomously (with various threshold criteria [i.e., any lethality, ⩾90% lethality, or 100% lethality]) or as a continuous variable (data not shown) and (2) lethality was bimodally distributed, with all strains killing either ⩽50% or ⩾90% of mice, strains yielding ⩾90% mouse lethality were classified as "killers."

Statistical methods. Comparisons of proportions were performed by Fisher's exact test, and comparisons of continuous variables were performed by Mann-Whitney *U* test. Each characteristic was coded as present or absent. The phylogenetic distribution of accessory traits was assessed by use of 4×2 tables, to examine each trait's prevalence across the 4 phylogenetic groups. Traits exhibiting a nonrandom distribution then underwent an assessment of prevalence within each phylogenetic group versus all other strains, in a series of 2×2 tables. Similarity relationships with respect to aggregate accessory-trait profiles were used to construct a dendrogram, by use of the unweighted-pair-group method with averaging (UPGMA) [24].

Logistic regression analysis (univariate and multivariate) to predict killer status was performed according to exact methods. Because many bacterial variables were closely correlated (data not shown), stepwise multivariate analysis was used. Because no statistical application was available for exact stepwise multivariate analysis, this function was approximated manually, as follows.

First, conventional stepwise multivariate analysis (by use of the SPSS program) was used to screen a candidate predictor-variable pool consisting of all significant univariate variables that did not have a zero cell. The criteria for entry into and retention within this model were $P < .05$ and $P < .10$, respectively; the last step of the model in which all variables yielded

Table 1. Distribution of bacterial characteristics, according to experimental virulence, among 90 *Escherichia coli* strains.

Category, specific trait[a]	Prevalence, no. (column %)					
		Killer (killed ⩾90% of mice)				
	Total (n = 90)	Yes (n = 37)	No (n = 53)	OR (95% CI)[b]	P	Nagelkerke R^2
Adhesins						
papA and *papEF*[c]	41 (46)	31 (84)	10 (19)	21.1 (6.6 to 80.3)	<.001	0.48
Allele F10 of *papA*	19 (21)	13 (35)	6 (11)	4.2 (1.3 to 15.2)	.01	0.11
Allele F11 of *papA*	11 (12)	10 (27)	1 (2)	18.7 (2.4 to 851.4)	<.001	0.19
papC	43 (48)	33 (89)	10 (19)	33.4 (9.2 to 160.3)	<.001	0.56
papG	33 (37)	24 (65)	9 (17)	8.7 (3.1 to 27.4)	<.001	0.29
Allele II	23 (26)	18 (49)	5 (9)	8.8 (2.7 to 35.0)	<.001	0.24
Allele III	17 (19)	15 (41)	2 (4)	16.8 (3.5 to 163.9)	<.001	0.27
sfa/focDE	24 (27)	20 (54)	4 (8)	13.9 (4.1 to 54.1)	<.001	0.33
sfaS	11 (12)	10 (27)	1 (2)	18.7 (2.4 to 851.4)	<.001	0.19
Toxins						
hlyD	27 (30)	22 (59)	5 (4)	13.6 (4.1 to 54.1)	<.001	0.35
cnf1	20 (22)	16 (43)	4 (8)	9.1 (2.5 to 41.8)	<.001	0.22
Siderophores						
iroN	43 (48)	27 (73)	16 (30)	6.1 (2.3 to 17.8)	<.001	0.23
fyuA	57 (63)	37 (100)	20 (34)	81.1 (13.3 to ∞)	<.001	0.56
ireA	21 (25)	13 (35)	8 (15)	3.0 (1.0 to 9.7)	.04	0.07
Capsule						
kpsM II	43 (48)	27 (73)	16 (30)	6.1 (2.2 to 17.8)	<.001	0.23
K1	17 (19)	14 (38)	3 (6)	9.9 (2.4 to 58.7)	<.001	0.21
Miscellaneous						
usp	47 (52)	34 (92)	13 (25)	33.1 (8.4 to 195.9)	<.001	0.53
ibeA	7 (8)	6 (16)	1 (2)	9.8 (1.1 to 471.6)	.02	0.09
ompT	58 (64)	35 (95)	20 (38)	22.1 (4.8 to 208.4)	<.001	0.37
H7 *fliC*	5 (6)	5 (14)	0 (0)	10.7 (1.4 to ∞)	.01	0.13
malX	49 (54)	35 (95)	14 (26)	46.1 (9.8 to 445.8)	<.001	0.55
Phylogenetic group						
Group A	31 (34)	1 (3)	30 (57)	.02 (0.001 to 0.2)	<.001	0.43
Group B2	47 (52)	33 (85)	14 (26)	22.0 (6.3 to 100.9)	<.001	0.47
Ecological source						
Fecal	18 (20)	1 (3)	17 (32)	0.06 (0.001 to 0.4)	<.001	0.20

[a] Characteristics listed in the table are those that yielded $P < .05$ (by Fisher's exact test) for prevalence among killers vs. nonkillers. Definitions are as follows: *cnf1,* cytotoxic necrotizing factor 1; *fyuA,* yersiniabactin system; *hlyD,* hemolysin; H7 *fliC,* flagellin variant; *ibeA,* invasion of brain endothelium A*; ireA* and *iroN,* siderophore receptors; *kpsM* II, group 2 capsules; K1, K1 capsule; *malX,* pathogenicity-island marker; *ompT,* outer-membrane protease; *pap,* P fimbriae (with *papA,* structural subunit and alleles thereof; *papEF,* tip pilins; *papC,* assembly; and *papG,* tip adhesin and alleles thereof); *sfa/focDE,* S and F1C fimbriae; *sfaS,* S fimbriae; and *usp,* uropathogenic-specific protein. Other characteristics detected in at least 1 strain each but not yielding $P < .05$ (and therefore not listed in the table) included (data in parentheses are % of 90 strains): *afa/draBC,* Dr-binding adhesins (6%); *afaE8, afa* variant (4%); *astA,* enteroaggregative *E. coli*–associated toxin (7%); *bmaE,* M fimbriae (6%); *cdtB,* cytolethal distending toxin (1%); *clpG,* adhesin (2%); *cvaC,* microcin V (11%); *focG,* F1C fimbriae (10%); *fimH,* type 1 fimbriae (100%); F17, adhesin (1%); *gafD,* G fimbriae (1%); phylogenetic groups B1 (4%) and D (9%); *iha,* adhesin-siderophore (31%); *iss,* increased serum survival (11%); *iutA,* aerobactin system (32%); *kpsMT* III, group 3 capsules (6%); K2, K2 capsule (6%); *papA* alleles F48 (8%), F14 and F536 (3% each), F7-1, F7-2, F8, and F12 (2% each), and F9, F13, F15, and F16 (1% each); *papG* allele I (1%); *rfc,* O4 lipopolysaccharide synthesis (3%); *sat,* secreted autotransporter toxin (38%); and *traT,* serum resistance associated (36%).

[b] CI, confidence interval; OR, odds ratio. Values are from exact univariate logistic regression analysis.

[c] *papA* and *papEF* gave identical results and therefore are shown together, for brevity.

$P < .05$ was used. Next, the significant predictor variables from this initial stepwise screen were entered into an exact multivariate logistic regression model (by use of SAS Proc Logistic). Then, significant univariate predictor variables that had a zero cell were added to the exact multivariate model, individually and jointly. These variables were retained in the model if (1) they yielded $P < .05$ in the presence of the other variables or (2) their addition caused another variable to lose significance ($P > .05$); in the latter instance, if the variable losing significance yielded $P > .10$, it was removed and the analysis was repeated. Additional models were created by repeating the entire process after significant predictors in previous models had been removed from the initial candidate predictor pool. The fraction of total variance accounted for by each model was inferred from the Nagelkerke R^2. Throughout, the criterion for statistical significance was $P < .05$.

Factorial analysis of correspondence (FAC) was used to describe associations among the various predictor variables, in relation to killer status [12, 25]. FAC uses a covariance matrix based on χ^2 distances. The computation determines a plane defined by 2 principal axes of the analysis; the first axis, F1, accounts for most of the variance, and the second axis, F2, which is orthogonal to F1, accounts for the largest part of the variance not accounted for by F1. FAC was performed by use of SPAD.N software (Cisia), with a 2-way table that had 90 rows (1 per *E. coli* strain) and 58 columns (1 per informative variable), including ecological source, phylogenetic group, mouse lethality, and accessory traits (table 1).

RESULTS

Mouse lethality. In the sc-infection model, 37 (41%) of the 90 *E. coli* strains killed either ⩾90% (11 strains) or 100% (26 strains) of the inoculated mice, establishing them as killers (figure 1). The remaining 53 (59%) strains either exhibited no mouse lethality (46 strains) or killed 10%–50% of inoculated mice (7 strains), establishing them as nonkillers. (Reclassifying or excluding the 7 strains that killed 10%–50% of mice did not appreciably alter the results [data not shown].)

Bacterial traits. Of the 90 *E. coli* strains, 52% belonged to group B2, 34% belonged to group A, 9% belonged to group D, and 4% belonged to group B1 (table 1). Each of the 55 accessory traits sought was detected in at least 1 isolate, at prevalences ranging from 100% (in the case of *fimH* [type 1 fimbriae]) to 1% (in the case of allele I of *papG,* 3 alleles of *papA, gafD* [G fimbriae], F17 [non-P adhesin], and *cdtB* [cytolethal distending toxin]) (table 1). A total of 25 accessory traits were significantly phylogenetically distributed. Of these, 23 (92%) were significantly concentrated within group B2, with or without a significantly decreased prevalence within group A and/or group B1 (table 2), whereas 1 (*sat*) was significantly concentrated within group A (at 54% prevalence, vs. 29% prevalence for the other accessory traits [$P = .02$]), and 1 (*astA*) was significantly concentrated within group B1 (at 50% prevalence, vs. 5% prevalence for the other accessory traits [$P = .02$]).

Univariate analysis of predictor variables versus lethality. A total of 23 variables—including 22 (40%) of the 55 accessory traits and group B2—were significantly positively associated with killer status, whereas 2 variables—namely, group A and fecal source—were significantly negatively associated with killer status (table 1). According to univariate logistic regression analysis, *papC, usp, malX,* and *fyuA* exhibited the highest odds ratios (ORs) (i.e., >25) favoring killer status and accounted for an estimated 53%–56% of total variance each (table 1); 5 other traits—*papA, papEF, ompT,* group B2, and group A—exhibited ORs >20 (or, inversely, <0.05) and accounted for an estimated 37%–48% of total variance each (table 1); and 6 other traits—allele III of *papG, sfa/focDE, sfaS, hlyD,* the K1 variant of *kpsM,* and *ibeA*—exhibited ORs >10 and accounted for an estimated 9%–27% of total variance each (table 1). Fecal source exhibited an OR of 0.06 for killer status and accounted for only 20% of total variance (table 1).

Multivariate models to predict killer status. The initial stepwise logistic regression analysis screen identified *papC, malX,* and *ireA* as the only significant predictors within the candidate predictor pool, which included all statistically significant univariate predictor variables (table 1) except (because of their zero cells) *fyuA* and H7. When *papC, malX,* and *ireA* were then entered into an exact multivariate logistic regression model, *papC* and *malX* appeared as strong positive predictors of killer status whereas *ireA* appeared as a moderately strong negative predictor of killer status, presumably because *papC* and *malX* accounted for all of *ireA*'s positive predictive power (table 3). This model (model 1a) accounted for an estimated

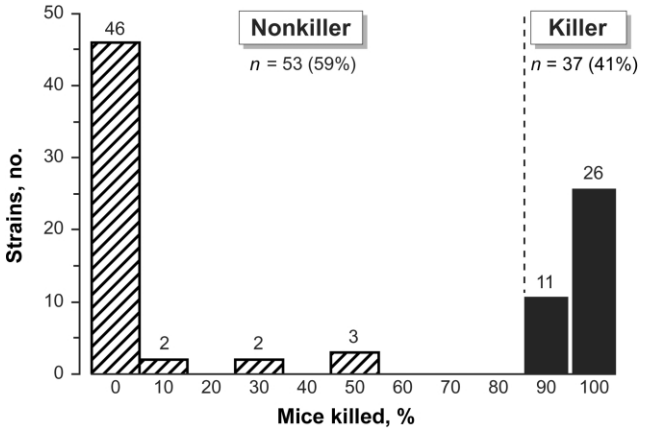

Figure 1. Distribution of 90 *Escherichia coli* strains, in terms of mouse lethality, in a standardized subcutaneous-infection model. Lethality in mice (10–20 mice/*E. coli* strain) was observed for 7 days postchallenge. On the basis of the distribution of the data and the results of pilot analyses, "Killer" status was assigned to strains yielding ⩾90% mouse lethality.

Table 2. Phylogenetic distribution of bacterial traits and source, among 90 *Escherichia coli* strains.

Category, specific trait[a]	Prevalence, no. (column %)			*P* for B2 vs. other groups[b]
		Phylogenetic group		
	Total (n = 90)	B2 (n = 47)	A, B1, or D (n = 43)	
Adhesins				
papA, papC, and *papEF*[c]	41–43 (46–48)	34–36 (72–74)	6–7 (14–16)	<.001
Allele F10 of *papA*	19 (21)	14 (30)	5 (17)	.04
Allele F11 of *papA*	11 (12)	11 (23)	0	.001
papG	33 (37)	26 (55)	7 (16)	<.001
Allele II	23 (26)	20 (43)	3 (7)	<.001
Allele III	17 (19)	16 (34)	1 (2)	<.001
sfa/focDE	24 (27)	24 (51)	0	<.001
sfaS	11 (12)	11 (23)	0	.001
focG	9 (10)	9 (19)	0	.003
Toxins				
hlyD	27 (30)	26 (55)	1 (2)	<.001
cnf1	20 (22)	19 (40)	1 (2)	<.001
sat	34 (38)	11 (23)	23 (53)	(.005)
Siderophores				
iroN	43 (48)	34 (72)	9 (21)	<.001
fyuA	57 (63)	43 (91)	14 (33)	<.001
ireA	21 (25)	19 (40)	2 (5)	<.001
Capsules				
kpsM II	43 (48)	33 (70)	10 (23)	<.001
K1	17 (19)	14 (30)	3 (7)	.007
Miscellaneous				
usp	47 (52)	44 (94)	3 (7)	<.001
ibeA	7 (8)	7 (15)	0	.01
ompT	58 (64)	44 (94)	14 (33)	<.001
malX	49 (54)	45 (96)	4 (9)	<.001
Ecological source				
Fecal	18 (20)	3 (6)	15 (35)	(.001)

[a] Characteristics listed in the table are the 23 that yielded $P < .05$ (by Fisher's exact test) for prevalence in group B2 vs. prevalence in all other strains combined, with or without a significantly decreased prevalence in 1 or more other phylogenetic groups. An additional 2 traits were significantly associated with group A (*sat*) or group B1 (*astA*), as described in the text. Definitions are as in table 1. Traits were assessed for their prevalence in each phylogenetic group vs. their prevalence in all other strains, if they were significantly phylogenetically distributed according to an initial 4×2-table screen involving all 4 phylogenetic groups.

[b] Parentheses denote a negative association with group B2.

[c] *papA, papC,* and *papEF* gave similar results and therefore are shown together, for brevity.

72% of total variance. When entered into this model, *fyuA* did not achieve statistical significance but did render *ireA* nonsignificant; the resulting model (model 1b) accounted for an estimated 75% of total variance. In contrast, when H7 was entered into the latter 2 models (i.e., those with and without *fyuA*), it neither achieved statistical significance nor rendered another variable nonsignificant (data not shown).

To assess the robustness of these models, additional models (models 2a–3c) were created after any variables identified as significant predictors in previous models had been removed from the initial candidate predictor pool. Thus, *papC, malX,* and *ireA* were removed from the candidate predictor pool, and the process described above was repeated. The new initial stepwise screen identified *usp* and *papEF* as the only 2 significant multivariate predictors. An exact multivariate model (model 2a) based on *usp* and *papEF* showed both of the latter to be

Table 3. Stepwise multivariate logistic regression analysis of bacterial traits as predictors of experimental mouse lethality, among 90 *Escherichia coli* strains.

Initial candidate predictors in screening pool[a]	Significant predictors[a]	Exact multivariate model, with and without *fyuA*				
		Model	Nagelkerke R^2	Variable[a,b]	OR (95% CI)	P[c]
As in table 1, except for *fyuA* and H7	*papC, malX,* and *ireA*	1a	0.72	*papC*	42.4 (4.8 to >999.9)	<.001
				malX	45.2 (47 to >999.9)	<.001
				ireA	0.08 (0.002 to 0.8)	(.03)[c]
		1b	0.75	*papC*	16.0 (1.5 to 885.1)	.02
				malX	30.3 (27.0 to >999.9)	.002
				ireA	0.1 (0.002 to 1.05)	(.06)[c]
				fyuA	4.2 (0.4 to ∞)	.25
Also except for *papC, malX,* and *ireA*	*usp* and *papEF*	2a	0.63	*usp*	13.8 (3.0 to 88.8)	<.001
				papEF	8.1 (2.0 to 35.6)	.002
		2b	0.69	*usp*	8.0 (1.5 to 56.0)	.009
				papEF	3.4 (0.7 to 17.1)	.14
				fyuA	7.1 (1.0 to ∞)	.49
		2c	0.66	*usp*	10.7 (2.2 to 72.9)	.001
				fyuA	23.5 (3.2 to ∞)	<.001
Also except for *usp* and *papEF*	*papA* and group B2	3a	0.59	*papA*	8.7 (2.3 to 36.7)	<.001
				group B2	8.6 (2.0 to 44.0)	.002
		3b	0.67	*papA*	3.1 (0.7 to 15.0)	.17
				group B2	5.4 (1.3 to 30.4)	.03
				fyuA	13.0 (1.6 to ∞)	.02
		3c	0.66	group B2	7.9 (1.8 to 42.4)	.004
				fyuA	30.7 (4.5 to ∞)	<.001

[a] Definitions are as follows: *fyuA,* yersiniabactin receptor; group B2, phylogenetic group B2; *ireA,* siderophore receptor; *malX,* pathogenicity island marker; *pap,* P fimbriae (with *papA,* structural subunit and alleles thereof; *papC,* assembly; and *papEF,* tip pilins); and *usp,* uropathogenic specific protein.

[b] The initial candidate predictor variable pool for the stepwise screen included all statistically significant univariate predictors (table 1), except for *fyuA* and H7 (flagellin variant). Variables listed in the table are those which emerged as statistically significant predictors of killer status in successive iterations of the stepwise screen, iterations that were performed after serial removal of previously identified significant predictors. The exact models were based on the significant variables from the stepwise screen, with or without *fyuA* (yersiniabactin receptor). H7 (flagellin variant) did not significantly alter the results when added into the exact models after the addition of *fyuA* (data not shown).

[c] Parentheses indicate negative association with killer status. Note that, although *ireA* exhibited a positive univariate association with killer status (table 1), it was a negative predictor in multivariate models 1a and 1b, which included *papC* and *malX*. (Only those *ireA*-positive strains that also were positive for *papC* and *malX* were killers [data not shown]).

strong positive predictors, together accounting for an estimated 63% of total variance. Addition of *fyuA* to this model rendered *papEF* nonsignificant, although *fyuA* itself also was nonsignificant. This model (model 2b) accounted for an estimated 69% of total variance. Removal of *papEF* (which was rendered nonsignificant after addition of *fyuA*) rendered *fyuA* highly significant and yielded a model (model 2c) that accounted for an estimated 66% of total variance. Addition of H7 to the latter 2 models (i.e., those with and without *fyuA*) had no significant effect (data not shown).

When this process was repeated after *usp* and *papEF* also had been removed from the initial candidate predictor pool, the stepwise screen identified *papA* and group B2 as the only 2 significant predictor variables (table 3). An exact multivariate model (model 3a) showed both *papA* and group B2 to be significant positive predictors and, jointly, to account for an estimated 59% of total variance. Addition of *fyuA* to this model rendered *papEF* nonsignificant, although *fyuA* itself also was nonsignificant; this model (model 3b) accounted for an estimated 69% of total variance. Removal of *papEF* (because of its nonsignificance after addition of *fyuA*) rendered *fyuA* highly significant and yielded a model (model 3c) that accounted for an estimated 66% of total variance. Addition of H7 to these 2 models (i.e., those with and without *fyuA*) had no significant effect (data not shown).

Cluster analysis. A dendrogram based on similarity relationships according to aggregate accessory-trait profiles contained 2 major clusters—cluster I and cluster II—that had 68% similarity and that comprised 40 and 50 strains, respectively (figure 2). Whereas cluster I contained predominantly killer strains (prevalence, 80%), cluster II was nearly devoid of killer strains (prevalence, 10%) ($P < .001$). Moreover, the few killer strains within cluster II were confined to a small subcluster, where they predominated (4/7 [57%], vs. 1/43 [2%] of other cluster II strains [$P < .001$]). Clusters I and II also differed significantly according to both ecological source (fecal: 5% and

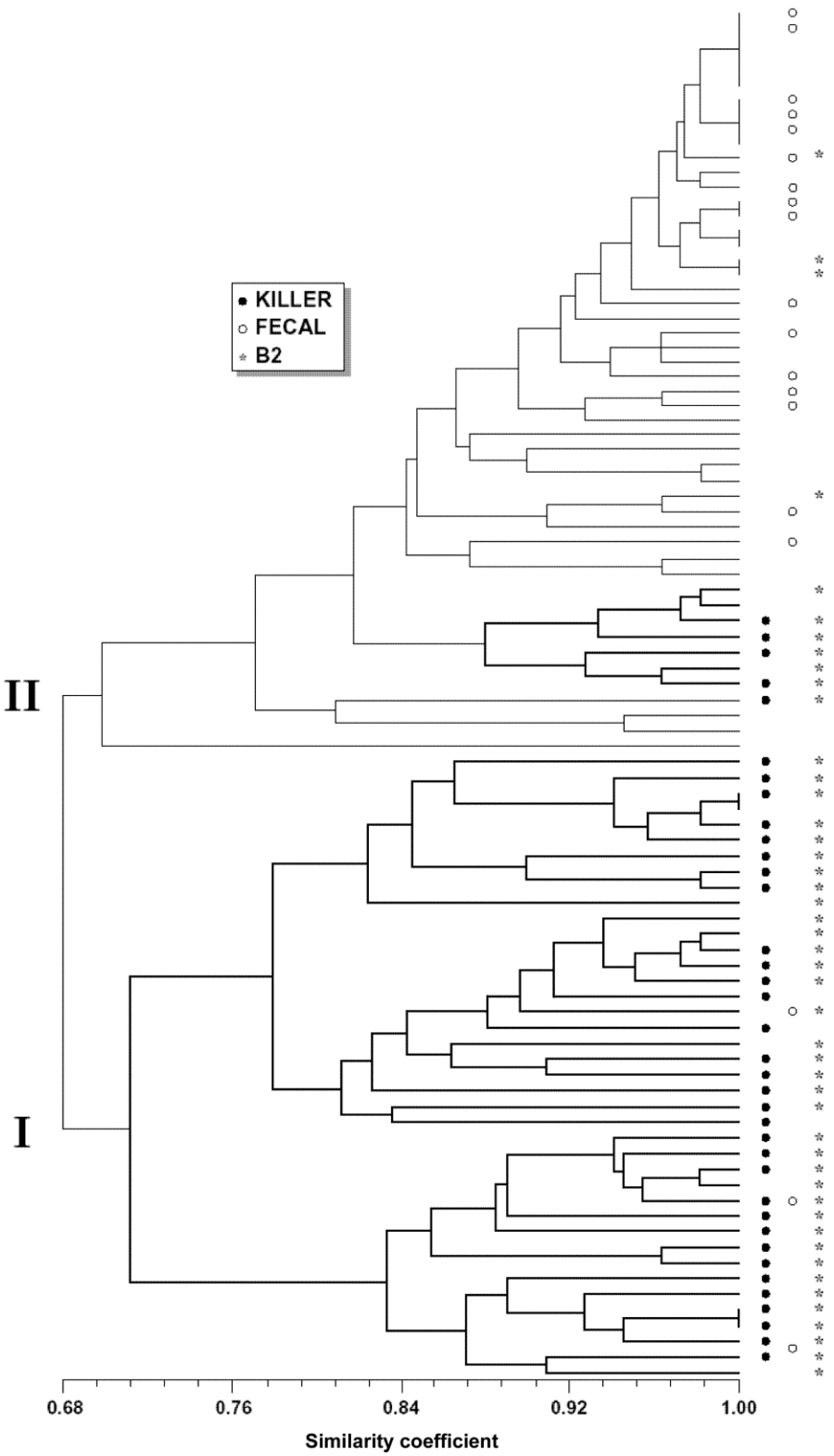

Figure 2. Dendrogram constructed by use of the unweighted-pair-group method with averaging and based on similarity relationships according to the presence or absence of 55 accessory traits among 90 *Escherichia coli* strains. Boldface lines indicate regions of the tree represented predominantly by "killer" isolates; the symbols to the right of the dendrogram denote each isolate's killer status (killer vs. nonkiller), ecological source (fecal vs. extraintestinal infection), and phylogenetic group of origin (group B2 vs. other).

32%, respectively [$P = .001$]) and phylogenetic origin (group B2: 93% and 18%, respectively [$P < .001$]) (figure 2); however, these differences between clusters I and II were less extreme than those for killer status (prevalence ratio: killer, 8.0; nonfecal, 6.4; group B2, 5.1).

FAC. To further explore the associations among bacterial characteristics and ecological source, in relation to killer status, FAC was performed. The F1/F2 plane accounted for 33.40% of total variance, with factors F1 and F2 accounting for 22.85% and 10.54%, respectively (figure 3). The planes obtained from the other factors of the FAC accounted for lower percentages of the variance than did F1 and F2 and did not further improve the data interpretation (data not shown).

The F1/F2 plane allowed the positioning of the variables according to their coordinates on each of these 2 factors (figure 3). The variable killer status was projected on the positive values of the 2 factors, with many of the other variables positioned around it; thus, 4 groups of variables—from those most to those least closely related to killer status—could be distinguished, by defining 4 zones of increasing diameter centered on killer status. The first group encompassed the accessory traits *papA, papC, papEF, papG, malX,* and *usp* (figure 3); phylogenetic group B2 also belonged to this group and, thus, was closely related to killer status and the principal killer-associated accessory traits. The second group encompassed K1, H7 *fliC, fyuA, iroN, hlyD, ibeA, sfa/focDE, focG, kpsM*II, *ompT, cnf1, ireA, rfc,* alleles F10, F11, F14, F15, and F48 of *papA,* and allele II of *papG.* The third group encompassed *afa/draBC, iutA, iha, iss, cvaC, traT, sfaS, gafD, clpG, afaE8, kpsMT* III, *bmaE,* K2, *cdtB,* alleles F7-1, F7-2, F8, F9, F12, F13, F16, and F536 of *papA,* alleles I and III of *papG,* and extraintestinal source. A fourth group, which was projected on the negative values of the F1 axis, encompassed *astA, sat,* F17, groups A, B1, and D, fecal source, and nonkiller status.

DISCUSSION

In this genotypic-phenotypic analysis of 90 fecal and extraintestinal-infection *E. coli* isolates, we found that many bacterial characteristics—including accessory traits (i.e. putative or proven "virulence factors"), phylogenetic background, and ecological source—significantly predicted experimental mouse lethality in an sc-infection model. Although the different accessory traits varied greatly as to the strength of their association with experimental virulence, several constituted the strongest overall predictors of virulence, superseding phylogenetic group and ecological source.

On the basis of univariate and multivariate analysis and FAC, accessory traits were stronger predictors of experimental mouse virulence than were phylogenetic group or ecological source.

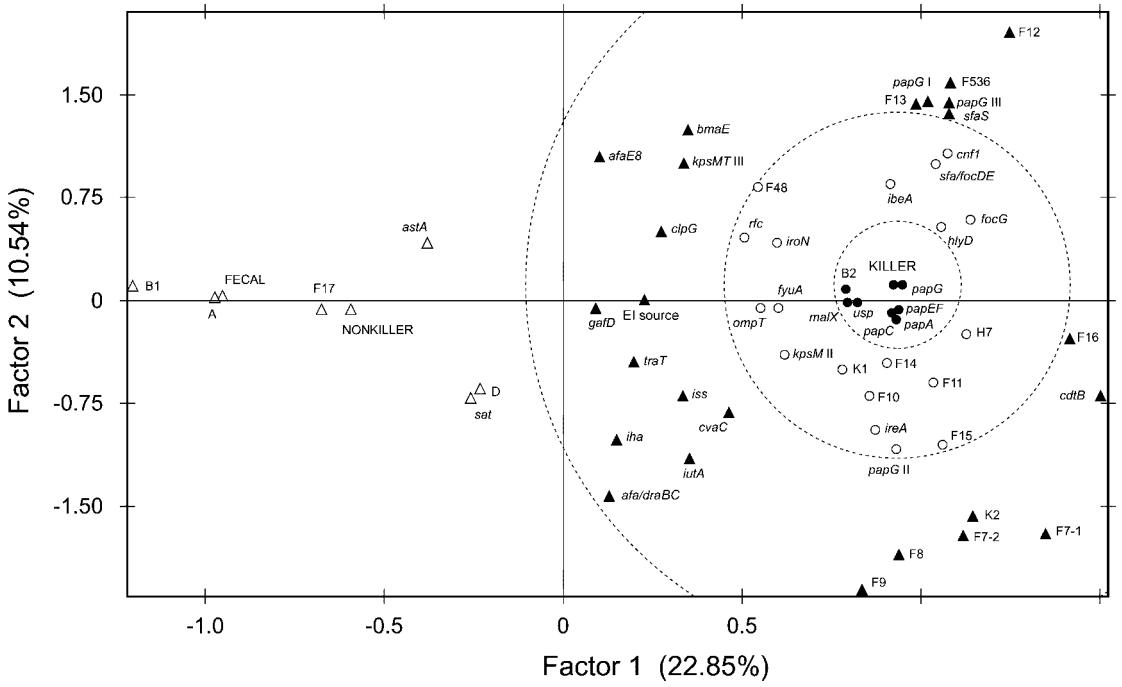

Figure 3. Factorial analysis of correspondence (FAC) between bacterial traits and mouse lethality, among 90 *Escherichia coli* strains. Projections on the F1/F2 plane, as computed by use of FAC, of 55 accessory traits (named as in table 1), phylogenetic group (A, B1, B2, and D), ecological source (fecal or infection), and killer vs. nonkiller status, are shown. The concentric circles *(dashed lines)* centered on the variable "KILLER" demarcate 4 groups of variables—group 1 (●); group 2 (○), group 3 (▲), and group 4 (△). The loading score for each variable, for factors F1 and F2, respectively, can be inferred from that variable's *X* and *Y* coordinates on the F1/F2 plane.

For example, in the multivariate models, only after 5 accessory traits (*papC, malX, ireA, usp,* and *papEF*) had been removed from the candidate predictor pool did group B2 appear as a significant predictor, and even then its predictive power was overshadowed by that of *fyuA*. Nonetheless, a close relationship was observed between group B2 and the accessory traits that best predicted virulence. These observations can be explained by the fact that, although most killer strains belong to group B2 and possess specific accessory traits, the few B2 strains that lack accessory traits do not kill mice, whereas the few non-B2 strains that have accessory traits can kill mice. This finding demonstrates that, although group B2 membership is not sufficient for virulence, it usually is associated with virulence. Indeed, virulence is a multifactorial condition, which is optimized by the association between the B2 phylogenetic background and specific accessory traits, as has been suggested by previous phylogenetic analyses showing the multiple independent arrivals of virulence factors into the B2 phylogenetic background [8].

At a practical level, the strong observed predictive power of the accessory traits indicates that molecular typing for such traits allows an excellent estimate of an individual *E. coli* isolate's intrinsic virulence potential, independent of ecological source, without the need for in vivo screening in an animal model. Alternatively, determination of phylogenetic group alone permits an approximate (albeit less accurate) estimate of virulence potential.

Many of the killer status–associated accessory traits—for example, *pap, sfa/foc, hly,* and *cnf*—represent traditionally recognized ExPEC virulence factors [26]. However, some traits—such as *iroN, fyuA, ireA, usp, ibeA, ompT,* and *malX*—have been identified relatively recently and/or have been studied comparatively little thus far [27–32], and these were absent from our previous study. Several of these traits were among the strongest overall predictors of virulence, and some have already been confirmed experimentally, by use of defined mutants, as virulence factors, although only in specific extraintestinal syndromes (e.g., neonatal meningitis, for *ibeA* [30]) or in specific hosts (e.g., poultry, for *malX* [33]). The findings of the present study support consideration of these traits as extraintestinal virulence factors of general relevance and indicate a need to explore the virulence role that they play in a broader range of syndromes and hosts, with attention to development of preventive measures such as vaccines [6, 34].

Consistent with previous observations [20], associations were noted—on the basis of FAC, for example—among the various accessory traits. Accordingly, in the multivariate analyses, little predictive power was lost by removing from the candidate predictor pool the significant variables in previous models, because closely correlated strong predictors entered the model to account for much the same variance as was accounted for by the removed variables. This finding reinforces the concept that virulence likely is multifactorial and that identification of specific traits as being highly predictive of virulence in a given multivariate model does not exclude the possible importance of other traits.

A unique advantage of the study design used is its ability to assess and compare multiple bacterial characteristics simultaneously as predictors of virulence, under standardized experimental conditions [13]. This capability is in contrast to the limitations of (1) gene-knockout studies, which typically address a single gene in an individual strain [9] and (2) molecular epidemiological studies, in which virulence is inferred (with uncertain validity) from clinical context [5]. Limitations of the study design include inherent uncertainty as to (1) whether the characteristics assessed participate directly in pathogenesis, rather than being simply markers, and (2) the applicability of the mouse sc-infection model to human infections at other extraintestinal sites. An additional limitation is the use of multiple comparisons, giving an increased probability of type I errors. (Notably, however, most *P* values were quite low, indicating that the associations were unlikely to represent chance findings.) Likewise, conclusions based on the multivariate logistic regression analyses must be tempered by the small sample size and the large number of potential predictor variables, which, taken together, can render estimates of ORs and their associated confidence intervals less precise; therefore, results of analyses based on the study design used should be regarded as exploratory/descriptive.

The present study greatly supersedes our previous, similar study [12]—with respect to the number of strains analyzed, the extent of the laboratory testing, and the complexity of the statistical analyses. Specifically, the present study involved 55 virulence traits (vs. 9) and prevalence comparisons (not previously done) of bacterial traits and source versus killer status, multivariate logistic regression analysis, and cluster analysis of extended virulence profiles. Accordingly, we were able to rank the accessory traits by their strength of association with virulence and to directly compare, as predictors of virulence, accessory traits versus phylogenetic background and ecological source. We thereby were able to reach numerous novel conclusions, including the supremacy, as predictors of virulence, of specific accessory traits (including several "newer" ones) and aggregate trait profiles versus phylogenetic background and ecological source.

In summary, we found that, in 90 *E. coli* strains, many group B2–associated accessory traits significantly predicted mouse lethality in a standardized sc-infection model, outperforming both phylogenetic group and ecological source. This finding provides further evidence that these (or closely linked) accessory traits play a major role in the enhanced virulence of group B2 strains. It also demonstrates that molecular typing both permits accurate predictions regarding the in vivo virulence

potential of an *E. coli* strain and identifies a number of promising potential targets for future preventive interventions against ExPEC infections.

Acknowledgments

Ann Emery (Minneapolis Veterans Administration Medical Center) helped with data entry and manuscript preparation. Brian Johnston (Minneapolis VA Medical Center) and Stéphanie Gouriou (Faculté de Médecine de Brest) provided invaluable technical assistance.

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. Jackson LA, Benson P, Neuzil KM, Grandjean M, Marino JL. Burden of community-onset *Escherichia coli* bacteremia in seniors. J Infect Dis **2005**; 191:1523–9.
3. Picard B, Journet-Mancy C, Picard-Pasquier N, Goullet P. Genetic structures of the B2 and B1 *Escherichia coli* strains responsible for extraintestinal infections. J Gen Microbiol **1993**; 139:3079–88.
4. Johnson JR. Microbial virulence determinants and the pathogenesis of urinary tract infection. Infect Dis Clin North Am **2003**; 17:261–78.
5. Johnson JR, Russo TA. Molecular epidemiology of extraintestinal pathogenic (uropathogenic) *Escherichia coli*. Int J Med Microbiol **2005**; 295: 383–404.
6. O'Hanley P. Prospects for urinary tract infection vaccines In: Mobley HLT, Warren JW, eds. Urinary tract infections: molecular pathogenesis and clinical management. Washington, DC: ASM Press, **1996**:405–25.
7. 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.
8. Escobar-Paramo P, Clermont O, Blanc-Potard A-B, Bui H, Le Bougenec C, Denamur E. A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. Mol Biol Evol **2004**; 21:1085–94.
9. Donnenberg MS. Microbial determinants of *E. coli* uropathogenicity [abstract 626]. In: 41st Interscience Conference on Antimicrobial Agents and Chemotherapy (Chicago, IL). Washington, DC: American Society for Microbiology, 2001:509.
10. 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.
11. Johnson JR, Moseley S, Roberts P, Stamm WE. Aerobactin and other virulence factor genes among strains of *Escherichia coli* causing urosepsis: association with patient characteristics. Infect Immun **1988**; 56: 405–12.
12. Picard B, Sevali Garcia J, Gouriou S, et al. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. Infect Immun **1999**; 67:546–53.
13. McClelland EE, Bernhardt P, Casadevall A. Estimating the relative contributions of virulence factors for pathogenic microbes. Infect Immun **2006**; 74:1500–4.
14. Clermont O, Bonacorsi S, Bingen E. Rapid and simple determination of the *Escherichia coli* phylogenetic group. Appl Environ Microbiol **2000**; 66:4555–8.
15. Hull RA, Gill RE, Hsu P, Minshew BH, Falkow S. Construction and expression of recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract infection *Escherichia coli* isolate. Infect Immun **1981**; 33:933–8.
16. Blum G, Ott M, Lischewski A, et al. Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. Infect Immun **1994**; 62:606–14.
17. Welch R, Burland V, Plunkett G III, et al. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli.* Proc Natl Acad Sci USA **2002**; 99:17020–4.
18. Badger JL, Wass CA, Weissman SJ, Kim KS. Application of signature-tagged mutagenesis for identification of *Escherichia coli* K1 genes that contribute to invasion of human brain microvascular endothelial cells. Infect Immun **2000**; 68:5056–61.
19. Blattner FR, Plunkett G, Bloch CA, et al. The complete genome sequence of *Escherichia coli* K-12. Science **1997**; 277:1453–74.
20. Johnson JR, Stell AL. Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. J Infect Dis **2000**; 181:261–72.
21. Johnson JR, Stell AL, Scheutz F, et al. Analysis of F antigen-specific *papA* alleles of extraintestinal pathogenic *Escherichia coli* using a novel multiplex PCR-based assay. Infect Immun **2000**; 68:1587–99.
22. Johnson JR, Gajewski A, Lesse AJ, Russo TA. Extraintestinal pathogenic *Escherichia coli* as a cause of invasive non-urinary infections. J Clin Microbiol **2003**; 41:5798–802.
23. Pitout JDD, Laupland KB, Church DL, Menard ML, Johnson JR. Virulence factors of *Escherichia coli* that produce CTX-M-type extended-spectrum b-lactamases. Antimicrob Agents Chemother **2005**; 49:4667–70.
24. Sokal RR, Sneath PHA. Construction of a taxonomic system In: Sokal RR, Sneath PHA, eds. Principles of numerical taxonomy. San Francisco: WH Freeman, **1963**:169–215.
25. Greenacre M. Correspondence analysis in medical research. Stat Methods Med Res **1992**; 1:97–117.
26. Johnson JR. Virulence factors in *Escherichia coli* urinary tract infection. Clin Microbiol Rev **1991**; 4:80–128.
27. Russo TA, McFadden CD, Carlino-MacDonald UB, Beanan JM, Barnard TJ, Johnson JR. IroN functions as a siderophore receptor and is a urovirulence factor in an extraintestinal pathogenic isolate of *Escherichia coli.* Infect Immun **2002**; 70:7156–60.
28. Russo TA, Carlino UB, Johnson JR. Identification of *ireA,* a novel iron regulated virulence gene in an extraintestinal pathogenic isolate of *Escherichia coli.* Infect Immun **2001**; 69:6209–16.
29. 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.
30. Kim KS. *Escherichia coli* translocation at the blood-brain barrier. Infect Immun **2001**; 69:5217–22.
31. Huang S-H, Wan ZS, Chen HM, Jong AY. OmpT contributing to *Escherichia coli* invasion of human endothelial cells [abstract B-58]. In: American Society for Microbiology 101st General Meeting (Orlando, FL). Washington, DC: American Society for Microbiology, **2001**:54.
32. Schubert S, Picard B, Gouriou S, Heesemann J, Denamur E. *Yersinia* high-pathogenicity island contributes to virulence in *Escherichia coli* causing extraintestinal infections. Infect Immun **2002**; 70:5335–7.
33. Li G, Laturnus C, Evers C, Wieler LH. Identification of genes required for avian *Escherichia coli* septicemia by signature-tagged mutagenesis. Infect Immun **2005**; 73:2818–27.
34. Russo TA, McFadden CD, Carlino-MacDonald UB, Beanan JM, Olson R, Wilding GE. The siderophore receptor IroN of extraintestinal pathogenic *Escherichia coli* is a potential vaccine candidate. Infect Immun **2003**; 71:7164–9.