JOURNAL OF CLINICAL MICROBIOLOGY, Dec. 2008, p. 3906–3911 Vol. 46, No. 12
0095-1137/08/$08.00+0 doi:10.1128/JCM.00949-08

Molecular Epidemiology and Phylogenetic Distribution of the *Escherichia coli pks* Genomic Island▽

James R. Johnson,[1,2]* Brian Johnston,[1,2] Michael A. Kuskowski,[3,4] Jean-Philippe Nougayrede,[5,6] and Eric Oswald[5,6]

Medical Service[1] and Geriatric Research, Education, and Clinical Center,[3] Veterans Affairs Medical Center, Minneapolis, Minnesota; Department of Medicine[2] and Department of Psychiatry,[4] University of Minnesota, Minneapolis, Minnesota; INRA, UMR 1225, F-31076 Toulouse, France[5]; and Université de Toulouse, ENVT, UMR 1225, F-31076 Toulouse, France[6]

Received 17 May 2008/Returned for modification 1 August 2008/Accepted 13 October 2008

Epidemiological and phylogenetic associations of the *pks* genomic island of extraintestinal pathogenic *Escherichia coli* (ExPEC), which encodes the genotoxin colibactin, are incompletely defined. *clbB* and *clbN* (as markers for the 5′ and 3′ regions of the *pks* island, respectively), *clbA* and *clbQ* (as supplemental *pks* island markers), and 12 other putative ExPEC virulence genes were newly sought by PCR among 131 published *E. coli* isolates from hospitalized veterans (62 blood isolates and 69 fecal isolates). Blood and fecal isolates and *clbB*-positive and -negative isolates were compared for 66 newly and previously assessed traits. Among the 14 newly sought traits, *clbB* and *clbN* (colibactin polyketide synthesis system), *hra* (heat-resistant agglutinin), and *vat* (vacuolating toxin) were significantly associated with bacteremia. *clbB* and *clbN* identified a subset within phylogenetic group B2 with extremely high virulence scores and a high proportion of blood isolates. However, by multivariable analysis, other traits were more predictive of blood source than *clbB* and *clbN* were; indeed, among the newly sought traits, only *pic* significantly predicted bacteremia (negative association). By correspondence analysis, *clbB* and *clbN* were closely associated with group B2 and multiple B2-associated traits; by principal coordinate analysis, *clbB* and *clbN* partitioned the data set better than did blood versus fecal source. Thus, the *pks* island was significantly associated with bacteremia, multiple ExPEC-associated virulence genes, and group B2, and within group B2, it identified an especially high-virulence subset. This extends previous work regarding the *pks* island and supports investigation of the colibactin system as a potential therapeutic target.

Escherichia coli, an important extraintestinal pathogen in humans and certain animals, is a common cause of urinary tract infection, sepsis, neonatal meningitis, and colibacillosis (12, 21). The specialized *E. coli* strains that cause most extraintestinal infections, often termed extraintestinal pathogenic *E. coli* (ExPEC), exhibit multiple accessory traits, such as adhesins, toxins, polysaccharide coatings, invasins, and siderophores, that are absent or uncommon among nonpathogenic commensal strains, such as *E. coli* K-12 (8, 12). These traits, commonly referred to as virulence factors (VFs), contribute to the distinctive ability of ExPEC to resist host defense mechanisms and injure or invade the host. A better understanding of ExPEC VFs could guide the development of preventive measures, such as vaccines against extraintestinal *E. coli* infections, and could be used in surveillance systems for identifying reservoirs and transmission of ExPEC strains (4, 9, 20).

We recently identified in *Escherichia coli* a hybrid nonribosomal peptide-polyketide compound called colibactin that induces DNA double-strand breaks in eukaryotic cells (16). Such DNA damage activates host cellular signaling pathways that lead to cell cycle arrest, the progressive enlargement of cell bodies, and eventually cell death. We identified a 54-kb genomic island that is responsible for this bacterial trait. This genomic island, which is hereafter referred to as the *pks* island, encodes the machinery for the synthesis of colibactin. This machinery consists of nonribosomal peptide megasynthases (NRPS), polyketide megasynthases (PKS), hybrid NRPS/PKS enzymes, different tailoring or editing enzymes, and a critical phosphopantetheinyl transferase that mediates NRPS and PKS activation (16). The *pks* island was first characterized in sequenced ExPEC prototype strains and up to now was found exclusively in phylogenetic group B2 strains (16). Many strains belonging to this phylogenetic group cause extraintestinal infections, such as urinary tract infections and septicemia. Colibactin can thus be regarded as a VF of ExPEC, but little is known about its prevalence among isolates from defined clinical syndromes versus commensal strains or its associations with other VFs (13, 16, 25). In addition, certain commensal strains from healthy people also possess the *pks* island (13). Of note, one such commensal, the probiotic strain Nissle 1917, contains the *pks* island and produces a functional genotoxin (16).

Accordingly, in the present study, we sought to better define the clinical correlates of this polyketide synthesis system and its associations with other ExPEC VFs and phylogenetic group. For this, we took advantage of a previously published collection of 62 bloodstream isolates from U.S. veterans with diverse-source *E. coli* bacteremia and 69 concurrent fecal isolates from uninfected veterans (22), for which the major *E. coli* phylogenetic group (A, B1, B2, and D) and 48 virulence genes

* Corresponding author. Mailing address: Infectious Diseases (111F), VA Medical Center, 1 Veterans Drive, Minneapolis, MN 55417. Phone: (612) 467-4185. Fax: (612) 727-5995. E-mail: johns007@umn.edu.

▽ Published ahead of print on 22 October 2008.

had already been analyzed (22). We newly characterized these isolates for *clbB* and *clbN* (as markers for the 5′ and 3′ regions of the *pks* island, respectively), *clbA* and *clbQ* (as supplemental *pks* island markers), and 12 other putative ExPEC virulence genes. We then compared *clbB* and *clbN* with the 60 other virulence genes and phylogenetic group as predictors of blood versus fecal source, examined the phylogenetic distribution of *clbB* and *clbN*, and defined the associations of *clbB* and *clbN* with other virulence genes.

MATERIALS AND METHODS

Patients and strains. The bacteremia and fecal isolates studied have been described previously (22). Briefly, all blood isolates of *E. coli* recovered by the Veterans Administration (VA) Medical Center clinical microbiology laboratory between 1 December 1996 and 30 September 1999 were obtained from the laboratory. For fecal controls, during the same time interval, newly admitted VA Medical Center inpatients who lacked evidence of acute infection were recruited as donors of rectal swabs. Patients identified to study personnel by their primary physicians or nurses as potential study subjects were approached regarding study participation. A rectal swab was collected from each consenting subject and processed as described elsewhere (3) to recover an isolated colony of presumptive *E. coli*, i.e., lactose- and indole-positive, citrate-negative, gram-negative bacilli with characteristic colonial morphology. Previous data show that an arbitrarily selected fecal *E. coli* colony has an 86% probability of representing the quantitatively predominant clone in the sample (14). Blood and fecal isolates were stored at −80 C in 15% glycerol pending molecular analysis.

Virulence genotyping. The 12 newly detected markers besides *clbB* and *clbN* (colibactin polyketide synthesis system) included *afaE8* (afimbrial adhesin variant), *astA* (enteroaggregative *E. coli* heat-stable toxin), *clpG* (K-88-related CS31A adhesin), the F17 gene (F17c fimbriae), *hlyF* (hemolysin variant), *hra* (heat-resistant agglutinin), *kfiC* (K5 capsule), *kpsM* K15 (K15 capsule), *pic* (protein involved in intestinal colonization; serine protease), *tsh* (temperature-sensitive hemagglutinin; serine protease), *usp* (uropathogen-specific protein), and *vat* (vacuolating autotransporter; serine protease). Primers for the F17, *sat*, *usp*, *afaE8*, *clpG*, *astA*, *kfiC*, and *pic* genes were as published previously (7, 17, 24). Primers for *clbB* and *clbN* were *clbB*r (r for reverse orientation) (5′-CCA TTT CCC GTT TGA GCA CAC-3′), *clbB*f (f for forward orientation) (5′-GAT TTG GAT ACT GGC GAT AAC CG-3′), *clbN*r (5′-CAG TTC GGG TAT GTG TGG AAG G-3′), and *clbN*f (5′-GTT TTG CTC GCC AGA TAG TCA TTC-3′). Primers for two internal loci within the *pks* island, i.e., *clbA* (phosphopantetheinyl transferase) and *clbQ* (thioesterase), were IHAPJPN42 (*clbA*r) (5′-CAG ATA CAC AGA TAC CAT TCA-3′), IHAPJPN46 (*clbA*f) (5′-CTA GAT TAT CCG TGG CGA TTC3′), IHAPJPN55 (*clbQ*r) (5′-TTA TCC TGT TAG CTT TCG TTC-3′), and IHAPJPN56 (*clbQ*f) (5′-CTT GTA TAG TTA CAC AAC TAT TTC-3′). Primers for the K15 *kpsM* variant and *tsh*, as designed based on published sequence data (15, 23), were *k15*f (5′-ACG GAT TCA CGA CAA AGC TC-3′), *k15*r (5′-GGC AAA TAT CGC TTG GGT TA-3′), *tsh*f (5′-GGC GCA TAT CTG GAT AAG GA-3′), and *tsh*r (5′-GCA GGT TAT CGC CCT TAA CA-3′). Testing was done in duplicate using separately prepared boiled lysates as template DNA, with appropriate positive and negative controls, and published amplification conditions (7, 10, 11).

Isolates were operationally regarded as ExPEC if they were positive for two or more of the following genes: *papA* and/or *papC*, *sfa* and/or *foc* (S and F1C fimbriae), *afa* and/or *dra* (Dr-binding adhesins), *iutA* (aerobactin system), and *kpsM* II (group 2 capsule). The virulence score was the number of virulence genes detected, adjusted for multiple detection of the *pap*, *sfa* or *foc*, and *kpsM* II operons.

Phylotyping. Major *E. coli* phylogenetic groups (A, B1, B2, and D) were determined by multiplex PCR. For several isolates, seeming discrepancies between newly determined virulence genotype and previously determined phylogenetic group (22) prompted repeat phylotyping, leading to their reclassification into a different phylogenetic group.

Statistical methods. Comparisons of proportions were tested using Fisher's exact test (two-tailed). Virulence scores were compared by using the Mann-Whitney U test. Odds ratios and 95% confidence intervals were calculated by using univariable logistic regression and stepwise multivariable logistic regression.

Similarity relationships among the individual isolates with respect to VF profiles and phylogenetic group were assessed by using principal coordinate analysis (PCoA), which is a multivariate technique related to correspondence analysis (2) that allows one to plot the major patterns within a multivariate data set, e.g., multiple loci and multiple samples (18). Using GenAlEx6 (18), PCoA was applied to the VF data set in such a way so as to collapse the multiple VFs for simplified among-group comparisons. Values for each isolate from the first three PCoA axes, which are the axes that capture most of the variance within the data set (18), were used in a one-way multivariate analysis of variance (MANOVA) to test for differences between resistant and susceptible isolates.

RESULTS

Virulence genotypes. The colibactin system markers *clbB* and *clbN* were detected in 58 (44%) of the 131 total isolates, always jointly. Additionally, in a subset of the isolates (40 isolates; 10 each *clbB* and *clbN* positive and 10 *clbB* and *clbN* negative for blood and fecal isolates alike), PCR-based detection of two additional *pks* island markers (*clbA* and *clbQ*) yielded results precisely concordant with those for *clbB* and *clbN*, implying the presence of a complete *pks* island in all *clbB*- or *clbN*-positive strains. Accordingly, *clbB* alone was used in subsequent statistical analyses.

clbB was significantly associated with bacteremia, occurring in over half of the bacteremia isolates but in fewer than one-third of fecal isolates (odds ratio, 2.96; $P = 0.003$), although it accounted for only 9% of the total bacteremia-versus-fecal isolate variance (Table 1). Many other bacterial traits were also associated with bacteremia; they exhibited odds ratios for bacteremia ranging from 2.33 (for *iroN* [siderophore receptor]) to 16.69 (*ompT* [outer membrane protease]) (median, 3.96) and individually accounted for 5% to 43% (median, 9%) of the bacteremia-versus-fecal variance (Table 1). Aside from *clbB*, among the significantly bacteremia-associated markers were two other of the newly studied genes, i.e., *hra* (heat-resistant agglutinin) and *vat* (vacuolating toxin), which exhibited odds ratios and r^2 values similar to those of *clbB* (Table 1). Thus, *clbB* was more closely associated with bacteremia than were most of the other virulence genes studied, but not as closely as some, exhibiting a mid-range odds ratio and r^2 value compared with other significantly bacteremia-associated genes. Consistent with the greater prevalence of many individual virulence genes (including *clbB* and *clbN*) in bacteremia isolates, they exhibited significantly higher aggregate virulence scores than did fecal isolates (median score of 13.5 [range, 0 to 21] versus 7.0 [range 0 to 18]; $P < 0.001$).

Multivariable analysis. Because of known associations among virulence genes and phylogenetic background, multivariable logistic regression analysis was used to identify independent predictors of bacteremia. All the virulence markers studied, plus the phylogenetic groups (A, B1, B2, and D), were used as candidate predictor variables in a forward stepwise model. The resulting model, which accounted for 65% of total bacteremia-associated variance, identified seven significant predictor variables. Of these, *ompT*, *iutA*, *papA*, and *sfa* and/or *foc* were positive predictors of bacteremia, whereas *pic*, group B2, and the F16 *papA* allele were negative predictors, after adjustment for the other variables in the model (Table 2). Thus, in this model, *clbB* was not a significant independent predictor, whereas the only newly identified predictor was *pic*, which exhibited a negative (adjusted) association with bacteremia.

ExPEC isolates. A greater proportion of bacteremia isolates than fecal isolates satisfied molecular criteria for ExPEC (82% versus 45%; $P < 0.001$). Among the 82 ExPEC isolates, five

TABLE 1. Bacterial traits significantly associated with blood source among 131 fecal or blood isolates of *Escherichia coli* from veterans[a]

Category and trait[b]	Prevalence of trait (no. [%]) in isolates:			*P* value (fecal vs blood isolates)	Odds ratio (fecal vs blood isolates)	95% Confidence interval	Negelkerke r^{2d}
	Total (n = 131)	Fecal (n = 69)	Blood (n = 62)				
Adhesins							
papA, *papC*, *papEF*, or *papG*	64 (49)	21 (30)	43 (69)	<0.001	5.17	2.46–10.89	0.19
F10 *papA* allele	26 (20)	6 (9)	20 (32)	0.001	5.0	1.85–13.49	0.12
papG allele II	42 (32)	12 (17)	30 (48)	<0.001	4.45	2.01–9.88	0.14
sfa or *focDE*	35 (27)	11 (16)	24 (39)	0.005	3.33	1.46–7.48	0.09
focG	17 (13)	4 (6)	13 (21)	0.017	4.31	1.32–14.04	0.07
iha	35 (27)	10 (14)	25 (40)	0.001	3.99	1.72–9.24	0.11
hra	**42 (32)**	**14 (20)**	**28 (45)**	**0.005**	**3.12**	**1.44–6.75**	**0.09**
Toxins							
hlyD	39 (30)	11 (16)	28 (45)	<0.001	4.34	1.92–9.82	0.13
cnf1	32 (24)	9 (13)	23 (37)	0.002	3.93	1.65–9.38	0.10
sat	37 (28)	13 (19)	24 (39)	0.019	2.67	1.21–5.90	0.06
clbB and clbN	**58 (44)**	**22 (32)**	**36 (58)**	**0.003**	**2.96**	**1.45–6.05**	**0.09**
vat	**78 (60)**	**34 (49)**	**44 (71)**	**0.013**	**2.52**	**1.22–5.19**	**0.06**
Siderophores							
iroN	46 (35)	18 (26)	28 (45)	0.028	2.33	1.12–4.86	0.05
fyuA	91 (69)	40 (56)	51 (82)	0.004	3.36	1.50–7.54	0.09
ireA	25 (25)	5 (7)	20 (32)	<0.001	6.10	2.12–17.50	0.13
iutA	48 (37)	14 (20)	34 (55)	<0.001	4.77	2.21–10.31	0.16
Capsule *kpsM* II	84 (64)	34 (49)	50 (81)	<0.001	4.29	1.95–9.42	0.14
Miscellaneous							
traT	70 (53)	30 (43)	40 (65)	0.02	2.36	1.17–4.78	0.06
ompT	66 (50)	15 (22)	51 (82)	<0.001	16.69	7.71–39.72	0.43
malX	80 (61)	36 (52)	44 (71)	0.02	2.24	1.09–4.62	0.05
ExPEC (combination of above)[c]	82 (63)	31 (45)	51 (82)	<0.001	5.68	2.54–12.72	0.19
Phylogenetic group B1	15 (11)	13 (19)	2 (3)	0.005	0.14	0.03–0.67	0.09

[a] A total of 36 genes or traits were detected in at least one isolate each but not yielding $P < 0.05$ (in comparisons of the percentage of positive isolates; newly detected traits shown in bold type): *papA* alleles F7-1 (5%), F7-2 (1.5%), F8 (<1%), F9 (<1%), F11 (14%), F12 (2%), F13 (5%), F14 (7%), F16 (3%); *papG* allele III, P-adhesin variant (15%); *sfaS*, S fimbriae (10%); *afa* and/or *draBC*, Dr-binding adhesins (5%); ***afaE8***, afimbrial adhesin variant (1.5%); *bmaE*, M fimbriae (1.5%); F17 fimbriae, mannose-resistant adhesin (1.5%); ***clpG***, mannose-resistant adhesin (<1%); ***tsh***, temperature-sensitive hemagglutinin (4%); *fimH*, type 1 fimbriae (97%); ***hlyF***, variant hemolysin (7%); *cdtB*, cytolethal distending toxin (5%); ***astA***, enteroaggregative *E. coli* toxin (3%); ***pic***, autotransporter protease (9%); K1, group 2 capsule variant (26%); K2, group 2 capsule variant (6%); ***kfiC***, K5 group 2 capsule variant (11%); *kpsM* III, group 3 capsule (1.5%); *rfc*, O4 lipopolysaccharide (4%); *cvaC*, microcin V (6%); *usp*, uropathogen-specific protein (62%); *ibeA*, invasion of brain endothelium (14%); *iss*, increased serum survival (8%); H7 *fliC*, flagellin variant (18%); and phylogenetic group A (11%), phylogenetic group B2 (61%), and phylogenetic group D (16%). Five traits were sought but not detected (newly sought traits shown in bold type): F15 and **F5360** *papA* allele, *papG* allele I, *gafD* (G fimbriae), and **K15** (group 2 capsule variant).

[b] The 25 individual traits shown are those that yielded $P < 0.05$ when the *P* values for fecal versus blood isolates were compared. Newly detected traits are shown in bold type. Definitions: *papA*, *papC*, *papEF*, and *papG*, P-fimbria structural subunit, assembly, tip pilins, and adhesin, respectively (all gave highly similar results; values shown are for *papA*); *sfa* and/or *focDE*, S and/or F1C fimbriae; *focG*, F1C fimbriae; *iha*, adhesin siderophore; ***hra***, heat-resistant agglutinin; *hlyD*, alpha-hemolysin; *cnf1*, cytotoxic necrotizing factor 1; *sat*, secreted autotransporter toxin; ***clbB***, colibactin (***clbN*** gave identical results); ***vat***, vacuolating toxin; *iroN*, salmochelin receptor; *fyuA*, yersiniabactin receptor; *ireA*, siderophore receptor; *iutA*, aerobactin receptor; *kpsM* II, group 2 capsule; *traT*, serum resistance associated; *ompT*, outer membrane protease; *malX*, pathogenicity island marker.

[c] ExPEC, extraintestinal pathogenic *E. coli*, which was operationally defined as the presence of at least two of the following genes: *papA* and/or *papC*, *sfa* and/or *focDE*, *afa* and/or *draBC*, *iutA*, and *kpsM* II.

[d] Negelkerke r^2, fraction of variance (blood versus fecal) accounted for by the indicated trait.

TABLE 2. Stepwise multivariable logistic regression analysis for predicting blood source among 131 fecal or blood isolates of *Escherichia coli* from veterans

Variable[a]	Step of entry in model	*P* value	Odds ratio	95% Confidence interval
ompT	1	<0.001	59.9	11.34–315.8
pic	**2**	**0.004**	**0.04**	**0.005–0.36**
iutA	3	0.008	6.9	1.7–28.6
Group B2	4	0.001	0.03	0.004–0.2
F16 *papA* allele	5	0.007	0.007	0.000–0.2
papA	6	0.017	5.1	1.3–19.6
sfa and/or *foc*	7	0.039	5.4	1.9–26.9

[a] Variables shown are those that yielded $P < 0.05$ when entered into the stepwise model along with previously entered traits. Definitions are as in Table 1, footnotes *a* and *b*. The newly detected trait (*pic*) is shown in bold type.

virulence genes differed significantly in prevalence according to clinical source. Of these, three were significantly more prevalent (*papEF* [P-fimbria tip pilins], *ireA* [siderophore receptor], and *ompT*) and two significantly less prevalent (*pic* and the H7 *fliC* [flagellin] variant) among bacteremia isolates than fecal isolates (Table 3), whereas *clbB* was similarly prevalent in each group (71%). Accordingly, ExPEC blood isolates exhibited significantly higher aggregate virulence scores than did ExPEC fecal isolates (median score of 14 [range, 5 to 21] versus 12 [range, 2.25 to 18]; $P = 0.01$).

Associations of *clbB* with other traits in relation to bacteremia. To identify possible explanations for the univariable but

TABLE 3. Bacterial traits significantly associated with blood source among 82 fecal or blood isolates of extraintestinal pathogenic *Escherichia coli* from veterans

Trait[a]	Prevalence of trait (no. [%]) in isolates:			*P* value (fecal vs blood isolates)
	Total (*n* = 82)	Fecal (*n* = 31)	Blood (*n* = 51)	
papEF	58 (71)	17 (55)	41 (80)	0.02
pic	**10 (12)**	**7 (23)**	**3 (6)**	**0.037**
ireA	24 (29)	5 (16)	19 (37)	0.048
ompT	58 (68)	9 (29)	47 (92)	<0.001
H7 *fliC* allele	24 (29)	14 (45)	10 (20)	0.02

[a] Note that traits shown are those that yielded $P < 0.05$. Definitions are as in Table 1, footnotes *a* and *b*. The newly detected trait (*pic*) is shown in bold type.

not multivariable association of *clbB* with bacteremia, associations of *clbB* with other virulence genes and phylogenetic group were examined. *clbB* was significantly associated with many of the virulence genes studied, including, among the newly sought markers, *hra*, *pic*, *vat*, and *kfiC* (Table 4). *clbB* was also confined to ExPEC isolates and to phylogenetic group B2, occurring in approximately 75% of either subset. Accordingly, aggregate virulence scores among *clbB*-positive isolates were significantly higher than the scores among *clbB*-negative isolates. This was true both in the total population overall and within relevant subgroups, including fecal isolates and blood isolates) and in group B2 (overall, fecal isolates, and blood isolates) (Table 5). Likewise, according to multiple linear regression, group B2 and *clbB* were each significantly associated with increased virulence scores when present together in the model ($P < 0.001$ for each). Additionally, within group B2, *clbB*-positive isolates occurred nearly twice as often among blood isolates (36/58 [62%]) as among *clbB*-negative isolates (8/24 [33%]) ($P = 0.006$). These findings suggested that *clbB* identified a particularly VF-rich and bacteremia-capable subset within group B2, with multiple traits other than (or in addition to) colibactin possibly contributing to the enhanced virulence.

PCoA. For simplified two-group comparisons and to avoid possible multiple-comparison artifacts resulting from collinearity among virulence genes and phylogenetic group, PCoA was used to assess the distribution of bacteremia and fecal isolates and, separately, the *clbB*-positive and *clbB*-negative isolates, according to all the bacterial traits studied. In a PCoA that included all virulence markers studied and the four phylogenetic groups, the initial MANOVA indicated that the first three PCoA axes, when considered jointly, significantly differentiated the bacteremia and fecal isolates ($P < 0.001$). When the three axes were assessed individually, axis 1 was found to account for only 16% ($P < 0.001$) and axis 3 for 5% ($P = 0.005$) of the total bacteremia versus fecal variance, whereas axis 2 was noncontributory. In contrast, in a PCoA that included the same variables except *clbB*, the initial MANOVA again indicated an overall differentiation of bacteremia from fecal isolates by the first three PCoA axes ($P < 0.001$). Individual assessment of each axis showed that axis 1 accounted for fully 59% ($P < 0.001$) of *clbB*-specific variance, whereas axes 2 and 3 were noncontributory. Thus, compared with clinical source, *clbB* was much better able to resolve distinct bacterial subpopulations according to the assessed characteristics.

TABLE 4. Association of *clbB* or *clbN* with other bacterial traits among 131 blood and fecal isolates of *Escherichia coli* from veterans

Category and trait[a]	Prevalence of trait according to *clbB* or *clbN* status (no. [%])		*P* value
	clbB and *clbN* negative (*n* = 73)	*clbB* and *clbN* positive (*n* = 58)	
Adhesins			
papA, *papC*, *papEF*, or *papG*	16 (22)	48 (83)	<0.001
F7-2 *papA* allele	0 (0)	7 (12)	0.003
F10 *papA* allele	3 (4)	23 (40)	<0.001
F11 *papA* allele	4 (5)	12 (24)	0.004
F14 *papA* allele	1 (1)	8 (14)	0.01
papG allele II	12 (16)	30 (52)	<0.001
papG allele III	2 (3)	18 (31)	<0.001
sfa and/or *focDE*	4 (5)	31 (53)	<0.001
sfaS	2 (2)	11 (19)	0.003
focG	3 (4)	14 (24)	0.001
iha	13 (18)	22 (38)	0.016
hra	**13 (18)**	**29 (50)**	**<0.001**
Toxins			
hlyD	7 (10)	32 (55)	<0.001
cnf1	3 (4)	29 (50)	<0.001
cdtB	0 (0)	7 (12)	0.003
sat	13 (18)	24 (41)	0.004
pic	**3 (4)**	**9 (16)**	**0.03**
vat	**20 (27)**	**58 (100)**	**<0.001**
Siderophores			
iroN	13 (18)	33 (57)	<0.001
fyuA	33 (45)	58 (100)	<0.001
ireA	3 (4)	22 (38)	<0.001
iutA	19 (26)	29 (50)	0.006
Capsule			
kpsM II	29 (40)	55 (95)	<0.001
K1 *kpsM* variant	11 (15)	23 (40)	0.002
***kfiC* (K5 capsule)**	**4 (5)**	**10 (17)**	**0.045**
Miscellaneous			
usp	23 (32)	58 (100)	<0.001
traT	33 (45)	37 (64)	0.037
ompT	23 (32)	43 (74)	<0.001
H7 *fliC* variant	3 (4)	21 (36)	<0.001
malX	23 (32)	57 (98)	<0.001
ExPEC (combination of above)	24 (33)	58 (100)	<0.001

[a] Note that the traits shown are those that yielded $P < 0.05$. Definitions are as in Table 1, footnotes *a* and *b*. The newly detected traits are shown in bold type.

DISCUSSION

In this molecular epidemiological analysis of blood and fecal *E. coli* isolates, we found that two of the recently discovered colibactin synthesis genes, *clbB* and *clbN*, were significantly associated with bacteremia, being present in 58% of blood isolates but only 32% of fecal isolates ($P = 0.003$). Additionally, they were significantly associated with multiple other virulence genes, were confined to phylogenetic group B2, and within group B2 occurred within an especially high-virulence subset that exhibited extremely elevated virulence scores and a significantly increased likelihood of causing bacteremia.

Notably, however, in univariable analyses, *clbB* and *clbN* were less closely associated with bacteremia than were certain other traditionally recognized virulence genes. Likewise, ac-

TABLE 5. Virulence scores according to *clbB* or *clbN* status among blood and fecal isolates of *Escherichia coli* from veterans

Population (total no. of isolates)	*clbB*- and *clbN*-positive isolates		*clbB*- and *clbN*-negative isolates		*P* value[a]
	No. of isolates in subgroup	Virulence score (median [range])	No. of isolates in subgroup	Virulence score (median [range])	
Total population					
Overall (131)	58	14.9 (8–21)	73	4 (0–15)	<0.001
Fecal (69)	22	13.9 (8–18)	47	2 (0–13.8)	<0.001
Blood (62)	36	15 (11–21)	26	8.5 (0–15)	<0.001
Group B2					
Overall (80)	58	14.9 (8–21)	22	10 (1–15)	<0.001
Fecal (38)	22	13.9 (8–18)	16	8.3 (1–13.8)	<0.001
Blood (42)	36	15 (11–21)	6	14 (12–15)	0.003

[a] *P* values for comparison of virulence scores were determined by the Mann-Whitney U test.

cording to multivariable logistic regression analysis, these markers were not a significant independent predictor of bacteremia, likely because they were overshadowed by more potent group B2-associated traits (i.e., *ompT*, *sfa* and/or *foc*, *papA*) (6) that in turn are associated with *clbB* or *clbN*. This leaves uncertainty as to whether the colibactin system per se contributes directly to virulence or instead is a marker for other traits that play a more direct role in pathogenesis. Experimental assessment of *clbB* or *clbN* mutants in an animal infection model are needed to answer this question.

In the PCoA, *clbB* explained a substantially greater proportion of total variance (according to bacterial traits) than did bacteremia versus fecal source, i.e., 59% (*clbB*) versus 16% (clinical source). This is probably because *clbB* is an excellent marker for group B2 and related virulence genes, which in turn were responsible for much of the variance in the molecular data set. In contrast, clinical source is significantly associated with, but somewhat removed from, these traits, with its effect being reduced by host status (which allows low-virulence strains to cause bacteremia in certain individuals) and the presence of ExPEC strains in the fecal flora of many uninfected individuals.

A similar statistical association between the *pks* island and sequences involved in adherence, iron acquisition, and lipopolysaccharide synthesis, as suggested by correspondence analysis, was noted by Le Gall et al. among group B2 isolates from diverse clinical sources and time periods (13). In that study, these same traits were associated with experimental mouse lethality, although primary associations and causality were not assessed. Similarly, in a molecular analysis of asymptomatic bacteriuria (ABU) *E. coli* isolates, Zdziarski et al. found *clb* sequences in five of six group B2 strains (including probiotic strain 83972), despite the absence of other typical B2-associated virulence traits, but in none of four non-B2 strains (25). This may indicate that the *pks* island contributes to sustained commensalism within the urinary tract in such group B2 strains. However, the small numbers, absence of a non-ABU comparison group, and uncertain functional status of the *pks* island in the strains leaves uncertainty regarding the implications of these findings.

In the present study, beside *clbB* and *clbN*, only 2 other markers among the 14 newly studied virulence genes were significantly associated with bacteremia in univariable analyses, and in the multivariable model, only 1 (*pic*) of the 14 genes was a significant predictor—and a negative one at that. This suggests that diminishing returns with respect to explaining bacteremia-versus-fecal source variance can be expected from adding more and more new virulence genes into molecular epidemiological surveys, unless markers that account for some of the currently unexplained variance can be discovered. The latter might apply particularly to non-B2-associated virulence traits, few of which are currently recognized, despite the not-inconsiderable representation of non-B2 isolates in clinical collections.

Our findings have several practical implications. First, on epidemiological grounds, i.e., the high prevalence of *clbB* and *clbN* among bacteremia isolates and the strong association of these genes with bacteremia, *clbB* and *clbN* appear to represent potential targets for preventive interventions, even if not the most strongly statistically supported (20). Experimental assessment is warranted. Second, other newly studied virulence genes (e.g., *hra* and *vat*) may also be potential candidate targets and hence also may deserve experimental assessment. Finally, no new "blockbuster" virulence markers were identified; moreover, no single virulence gene accounted for more than 45% of blood-fecal variance and only one accounted for more than 19%. This supports the notion that extraintestinal virulence in *E. coli* is multiply determined, i.e., likely involves not only a constellation of accessory traits (i.e., virulence factors) but also possibly metabolic functions and other phenotypes not traditionally considered virulence related (1, 5, 19).

The present study confirms previously demonstrated epidemiological association of *clbB* and *clbN* with clinical (as opposed to fecal) isolates and with group B2 (13, 16). Furthermore, it adds significantly to previous work by studying a well-defined clinical population, i.e., bacteremia isolates, by determining the associations of *clbB* and *clbN* with multiple other traditional and recently recognized virulence genes and by using multiple complementary analytical approaches to establish the relationships among these bacterial traits, including phylogenetic group and clinical source. Limitations include the single study locale, predominantly elderly male veteran study population, the modest sample size (with its risk of type 2 errors, although the many observed significant associations reduce this concern), use of multiple comparisons (with its risk of type 1 errors, which were guarded against by using PCoA and virulence scores), and reliance on presence/absence testing. Future study of isolates from different locales, syndromes, host groups, and time periods is needed to round out our understanding of the epidemiology of *clbB* and *clbN*.

In summary, we found that the recently discovered *E. coli* colibactin synthesis genes, *clbB* and *clbN*, (i) were significantly associated with bacteremia and with multiple other virulence genes, (ii) were confined to phylogenetic group B2, and (iii) within group B2 identified an especially high-virulence subset that exhibited extremely elevated virulence scores and an increased likelihood of causing bacteremia. These findings confirm and extend previous work regarding *clbB* and *clbN* and support investigation of the colibactin system as a potential target for preventive or therapeutic measures.

ACKNOWLEDGMENTS

This material is based upon work supported by the Office of Research and Development, Medical Research Service, Department of Veterans Affairs (J.R.J. and M.A.K.), grants from the National Institutes of Health (J.R.J.), and a grant from the ERA-NET program "Pathogenomic" (E.O. and J.-P.N.).

REFERENCES

1. **Billips, B. K., S. G. Forrestal, M. T. Rycyk, J. R. Johnson, D. J. Klumpp, and A. J. Schaeffer.** 2007. Modulation of host innate immune response in the bladder by uropathogenic *Escherichia coli*. Infect. Immun. **75:**5353–5360.
2. **Greenacre, M.** 1992. Correspondence analysis in medical research. Stat. Methods Med. Res. **1:**97–117.
3. **Johnson, J. R., and J. J. Brown.** 1998. Colonization with and acquisition of uropathogenic *Escherichia coli* strains as revealed by polymerase chain reaction-based detection. J. Infect. Dis. **177:**1120–1124.
4. **Johnson, J. R., and C. Clabots.** 2006. Sharing of virulent *Escherichia coli* clones among household members of a woman with acute cystitis. Clin. Infect. Dis. **43:**e101–e108.
5. **Johnson, J. R., O. Clermont, M. Menard, M. A. Kuskowski, B. Picard, and E. Denamur.** 2006. Experimental mouse lethality of *Escherichia coli* isolates, in relation to accessory traits, phylogenetic group, and ecological source. J. Infect. Dis. **194:**1141–1150.
6. **Johnson, J. R., P. Delavari, M. Kuskowski, and A. L. Stell.** 2001. Phylogenetic distribution of extraintestinal virulence-associated traits in *Escherichia coli*. J. Infect. Dis. **183:**78–88.
7. **Johnson, J. R., A. Gajewski, A. J. Lesse, and T. A. Russo.** 2003. Extraintestinal pathogenic *Escherichia coli* as a cause of invasive nonurinary infections. J. Clin. Microbiol. **41:**5798–5802.
8. **Johnson, J. R., and T. A. Russo.** 2005. Molecular epidemiology of extraintestinal pathogenic (uropathogenic) *Escherichia coli*. Int. J. Med. Microbiol. **295:**383–404.
9. **Johnson, J. R., M. R. Sannes, C. Croy, B. Johnston, C. Clabots, M. A. Kuskowski, J. Bender, K. E. Smith, P. L. Winokur, and E. A. Belongia.** 2007. Antimicrobial drug-resistant *Escherichia coli* isolates from humans and poultry products, Minnesota and Wisconsin, 2002–2004. Emerg. Infect. Dis. **13:**838–846.
10. **Johnson, J. R., and A. L. Stell.** 2000. Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. J. Infect. Dis. **181:**261–272.
11. **Johnson, J. R., A. L. Stell, F. Scheutz, T. T. O'Bryan, T. A. Russo, U. B. Carlino, C. C. Fasching, J. Kavle, L. van Dijk, and W. Gaastra.** 2000. Analysis of F antigen-specific *papA* alleles of extraintestinal pathogenic *Escherichia coli* using a novel multiplex PCR-based assay. Infect. Immun. **68:**1587–1599.
12. **Kariyawasam, S., T. J. Johnson, C. Debroy, and L. K. Nolan.** 2006. Occurrence of pathogenicity island I (APEC-O1) genes among *Escherichia coli* implicated in avian colibacillosis. Avian Dis. **50:**405–410.
13. **Le Gall, T., O. Clermont, S. Gouriou, B. Picard, X. Nassif, E. Denamur, and O. Tenaillon.** 2007. Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. Mol. Biol. Evol. **24:**2373–2384.
14. **Lidin-Janson, G., B. Kaijser, K. Lincoln, S. Olling, and H. Wedel.** 1978. The homogeneity of the faecal coliform flora of normal school-girls, characterized by serological and biochemical properties. Med. Microbiol. Immunol. **164:**247–253.
15. **Maurer, J. J., T. P. Brown, W. L. Steffens, and S. G. Thayer.** 1998. The occurrence of ambient temperature-regulated adhesins, curli, and the temperature-sensitive hemagglutinin Tsh among avian *Escherichia coli*. Avian Dis. **42:**106–118.
16. **Nougayrede, J. P., S. Homburg, F. Taieb, M. Boury, E. Brzuszkiewicz, G. Gottschalk, C. Buchrieser, J. Hacker, U. Dobrindt, and E. Oswald.** 2006. *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. Science **313:**848–851.
17. **Nowrouzian, F., I. Adlerberth, and A. E. Wold.** 2001. P fimbriae, capsule and aerobactin characterize colonic resident *Escherichia coli*. Epidemiol. Infect. **126:**11–18.
18. **Peakall, R., and P. E. Smouse.** 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes **6:**288–295.
19. **Russo, T. A., S. T. Jodush, J. J. Brown, and J. R. Johnson.** 1996. Identification of two previously unrecognized genes (*guaA*, *argC*) important for uropathogenesis. Mol. Microbiol. **22:**217–229.
20. **Russo, T. A., and J. R. Johnson.** 2006. Extraintestinal isolates of *Escherichia coli*: identification and prospects for vaccine development. Expert Rev. Vaccines **5:**45–54.
21. **Russo, T. A., and J. R. Johnson.** 2003. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: an overlooked epidemic. Microbes Infect. **5:**449–456.
22. **Sannes, M. R., M. A. Kuskowski, K. Owens, A. Gajewski, and J. R. Johnson.** 2004. Virulence factor profiles and phylogenetic background of *Escherichia coli* isolates from veterans with bacteremia versus uninfected control patients. J. Infect. Dis. **190:**2121–2128.
23. **Schneider, G., U. Dobrindt, H. Bruggemann, G. Nagy, B. Janke, G. Blum-Oehler, C. Buchrieser, G. Gottschalk, L. Emody, and J. Hacker.** 2004. The pathogenicity island-associated K15 capsule determinant exhibits a novel genetic structure and correlates with virulence in uropathogenic *Escherichia coli* strain 536. Infect. Immun. **72:**5993–6001.
24. **Wallace-Gadsden, F., J. R. Johnson, J. Wain, and I. N. Okeke.** 2007. Enteroaggregative *Escherichia coli* related to uropathogenic clonal group A. Emerg. Infect. Dis. **13:**757–760.
25. **Zdziarski, J., C. Svanborg, B. Wullt, J. Hacker, and U. Dobrindt.** 2008. Molecular basis of commensalism in the urinary tract: low virulence or virulence attenuation? Infect. Immun. **76:**695–703.