

Escherichia coli Strains from Pregnant Women and Neonates: Intraspecies Genetic Distribution and Prevalence of Virulence Factors

Stéphane Watt,^{1*} Philippe Lanotte,¹ Laurent Mereghetti,¹ Maryvonne Moulin-Schouleur,² Bertrand Picard,³ and Roland Quentin¹

Département de Microbiologie Médicale et Moléculaire, EA 3250, CHU Bretonneau, 37044 Tours Cedex,¹ INRA, Centre de Recherche de Tours, URO 86, 37380 Nouzilly,² and Département de Microbiologie, CHU Morvan, 29609 Brest,³ France

Received 2 December 2002/Returned for modification 9 January 2003/Accepted 14 February 2003

To determine the extent to which the vagina, endocervix, and amniotic fluid screen the *Escherichia coli* strains responsible for neonatal infections, we studied the genetic relationships among 105 *E. coli* strains isolated from all of the ecosystems involved in this infectious process. Twenty-four strains were isolated from the intestinal flora, and 25 strains were isolated from the vaginas of pregnant women. Twenty-seven strains were isolated from the amniotic fluid, blood, and cerebrospinal fluid (CSF) of infected neonates. The intraspecies genetic characteristics of all of the isolates were determined by random amplified polymorphic DNA (RAPD) analysis, PCR ECOR (*E. coli* reference) grouping, and PCR virulence genotyping. A correlation was found between the intraspecies distributions of the strains in the A, B1, B2, and D ECOR groups and in the two major RAPD groups (I and II). Nevertheless, the distribution of the *E. coli* strains in the RAPD groups according to their anatomical origins was more significant than their distribution in the ECOR groups. This may be explained by the existence of an *E. coli* subpopulation, defined by the RAPD I group, within the ECOR B2 group. This RAPD I group presents a major risk for neonates: 75% of the strains isolated from patients with meningitis and 100% of the strains isolated from patients with bacteremia were in this group. The vagina and the amniotic fluid are two barriers that favor colonization by highly infectious strains. Indeed, only 17% of fecal strains belonged to the RAPD I group, whereas 52% of vaginal strains and 67% of amniotic fluid strains belonged to this subpopulation. The *ibeA* and *iucC* genes were significantly associated with CSF strains, whereas the *hly* and *sfa/foc* genes were more frequent in blood strains. These findings could serve as a basis for developing tools to recognize vaginal strains, which present a high risk for neonates, for use in prophylaxis programs.

Escherichia coli and *Streptococcus agalactiae* are the major causes of neonatal meningitis and septicemia. These diseases are still major health problems in industrialized countries. *E. coli*, a typical inhabitant of the intestinal flora and a commensal of the vaginal flora, is also implicated in a large number of diseases. The incidence of neonatal *E. coli* infection is estimated to be between 0.2 and 5 per 1,000 live births (4). Neonatal colonization often results from maternal transmission during delivery (30). Thus, vaginal colonization, observed in 3 to 20% of pregnant women, seems to be an important step in neonatal infection. About half of all vaginal strains express the K1 antigen (1, 17), the presence of which has been associated with neonatal meningitis. However, only some of these women give birth to an infected child or to a very low birth weight child or give birth prematurely (17). This implies that only a few *E. coli* strains can complicate pregnancy or cause neonatal infections.

The *E. coli* population has a clonal organization. Phylogenetic studies of the 72 strains of the *E. coli* reference (ECOR) collection divided the species into four main phylogenetic groups (ECOR A, B1, B2, and D) (8, 10, 22). Neonatal sepsis

and meningitis are associated with a limited number of phylogenetic groups, even though a large number of *E. coli* strains have been isolated from the human fecal flora (3, 9, 21, 25, 31). In addition, neonatal pathogenicity due to *E. coli* is correlated with the presence of various virulence factors (VFs), but none of the pathogenic determinants studied are sufficient by themselves to allow the detection of strains able to cause neonatal meningitis or septicemia (2, 3, 13, 26).

E. coli strains involved in neonatal infections are thought to originate from the natural flora of pregnant women. Thus, to infect neonates, *E. coli* strains have to adapt to various ecological media. The first are the peculiar physicochemical conditions of the vaginal cavity, which differ considerably from those of the intestinal tract. Secondly, the bacteria have to cross the endocervix and survive in the amniotic fluid. Finally, the bacteria may be subjected to strong selection pressure by these conditions before generating neonatal septicemia and meningitis. The objective of this study was to determine the extent to which the vagina, endocervix, and amniotic fluid screen strains that can infect neonates. For this, we studied the genetic link among epidemiologically unrelated *E. coli* strains isolated from all relevant anatomical sites located between the digestive tract of pregnant women and the central nervous system of neonates. The relationship between isolates was determined by studying a number of genetic features, especially intraspecies genetic characteristics and the prevalence of viru-

* Corresponding author. Mailing address: Département de Microbiologie Médicale et Moléculaire EA 3250 CHU Bretonneau, 2 Blvd. Tonnellé, 37044 TOURS Cedex 01, France. Phone: (33) 2 47 47 80 56. Fax: (33) 2 47 47 38 12. E-mail: watt@med.univ-tours.fr.

lence genes. A better understanding of the isolates located at each anatomical level of the genital tract of pregnant women is important for the implementation of appropriate screening and prophylaxis programs with the aim of reducing the number of neonatal infections.

MATERIALS AND METHODS

Bacterial isolates and DNA preparation. A total of 105 epidemiologically unrelated *E. coli* strains were collected between 1989 and 1997 in 12 general hospitals throughout France (Aix en Provence, Annemasse, Avranches, Brest, Limoges, Lyon, Martigues, Quimper, Saint Briec, Tours, Vannes, and Vienne). Each isolate was identified by a biochemical method (Api 20E; bioMérieux, Marcy l'Etoile, France). Twenty-four strains were isolated from the fecal flora of asymptomatic pregnant women. All of the women were informed of the objective of the study during routine prenatal checkups. Women who agreed to participate were asked to give a fecal sample for bacterial culture. Twenty-five strains were isolated from the vaginal flora of asymptomatic pregnant women. Vaginal samples were collected in accordance with the French recommendations concerning universal prenatal screening for *S. agalactiae* colonization (Agence Nationale d'Accréditation et d'Évaluation en Santé [http://www.anaes.fr]). The other 56 strains were isolated during infections: 27 were isolated from the amniotic fluid collected at birth from asymptomatic neonates born in a chorioamnionitis context, 17 were isolated from the blood of newborns without meningitis, and 12 were isolated from the cerebrospinal fluid (CSF) of neonates.

Each strain was subcultured on horse blood trypticase soy agar plates for 24 h at 37°C. The culture was harvested in 1 ml of Tris-EDTA buffer and lysed by adding 150 µl of lysozyme (1%; Sigma, St. Louis, Mo.). After 30 min at 37°C, 25 µl of 1% pronase (Calbiochem, San Diego, Calif.) and 150 µl of 10% sodium dodecyl sulfate (Bio-Rad, Hercules, Calif.) were added, and the mixture was incubated at 56°C for 1 h. DNA was extracted by the phenol-chloroform-isoamyl alcohol (25/24/1) procedure and stored at 4°C.

Analysis of intraspecies genetic structure. The phylogenetic relationships among the 105 *E. coli* strains were studied by two methods: (i) PCR ECOR grouping (7) and (ii) the random amplified polymorphic DNA (RAPD) method, which is a highly valuable molecular tool for intraspecies characterization of *E. coli* (5, 8, 32).

PCR ECOR grouping. The strains were classified according to the ECOR system (10) by use of the rapid phylogenetic grouping technique described by Clermont et al. (7). This method is based on a triplex PCR involving the amplification of two genes (*chuA* and *yjaA*) and of an anonymous fragment of DNA from *E. coli*. Briefly, a two-step PCR was performed with three primer pairs: *chuA*1 (5'-GACGAACCAACGGTCAGGAT-3') and *chuA*2 (5'-TGCCGCCA GTACCAAAAGACA-3'), *yjaA*1 (5'-TGAAGTGTCAGGAGACGCTG-3') and *yjaA*2 (5'-ATGGAGAATGCGTTCCTCAAC-3'), and TspE4C2.1 (5'-GAGTA ATGTCGGGGCATTCA-3') and TspE4C2.2 (5'-CGCGCAACAAAGTATT ACG-3'). The results were interpreted as follows, according to the method of Clermont et al. (7): *chuA* positive and *yjaA* positive, group B2; *chuA* positive and *yjaA* negative, group D; *chuA* negative and TspE4C2 positive, group B1; *chuA* negative and TspE4C2 negative, group A.

RAPD fingerprinting. RAPD analysis was performed as described by Williams et al. (33) with some modifications. The 25-µl PCR mixture consisted of 100 µM (each) deoxyribonucleoside triphosphate (Boehringer), 0.2 µM (each) primer, 25 ng of DNA, and 0.5 U of *Taq* DNA polymerase (Perkin-Elmer, Norwalk, Conn.) in 1× buffer. The two primers (188, 5'-ACCTGGACCAAGGAGAAAAC-3', and 3388, 5'-ACCGGCCCGTACT-3') were chosen from among those proposed by Desjardin et al. (8) and synthesized by Eurogentec (Seraing, Belgium). Amplification was performed in a DNA Thermal Cycler 9600 (Perkin-Elmer) as described by Cave et al. (5). The amplification products were separated by electrophoresis in a 1.5% agarose gel and visualized by ethidium bromide staining. A 1-kb DNA ladder (Life Technologies, Gaithersburg, Md.) was used in each gel as a molecular size marker. A negative control, consisting of the same mixture but with water instead of DNA, was included in each run. This procedure was repeated twice for each strain. Only fragments present in both amplifications were used to define the RAPD patterns. RAPD patterns obtained with primers 188 and 3388 were digitized with a video camera connected to a microcomputer (Bio-Profil; Vibert-Lourmat, Marne la Vallée, France). The numerical analysis was carried out with the Taxotron software package (Taxolab: Institut Pasteur, Paris, France), which includes four programs (RestrictoScan, RestrictoTyper, Adanson, and Dendrograf), as previously described (6). For each strain, the RAPD type was defined as the combination of the patterns obtained with both primers. The relationships among the RAPD types of the strains were calculated

TABLE 1. Nucleotide sequences of PCR primers used to amplify eight virulence factors^a

Gene	Primer sequence (5'-3')	T _m (°C) ^b	Amplicon size (bp)	Reference
<i>hly</i>	AACAAGGATAAGCACTGTTCTGGCT ACCATATAAGCGGTTCATTCCTCGTCA	63	1,177	34
<i>papC</i>	GACGGTGTACTGCGAGGTGTGGCG ATATCCTTTCTGCAGGGATGCAATA	65	328	18
<i>sfa/foc</i>	CTCCGAGAAGCTGGGTGCATCTTAC CGGAGGAGTAATTACAAACCTGGCA	65	410	18
<i>afa</i>	GCTGGGACGCAAACTGATAACTCTC CATCAAGCTGTTTGTTCGTCGCCCG	65	750	18
<i>ibeA</i>	TTACC GCCGTGATGTTATCA CATTAGCTCTCGTTACCGCT	60	171	2
<i>iucC</i>	AAACCTGGCTTACGCAACTGT ACCCGCTGCAAATCATGGAT	55	269	2
<i>neuC</i>	AGGTGAAAAGCCTGGTAGTGTG GGTGGTACATTCGGGATGTG	61	675	35
<i>fimH</i>	TGCAGAACGGATAAGCCGTGG GCAGTCACTGCCCTCCGGTA	55	508	14

^a Alpha-hemolysin (*hly*), P fimbriae (*papC*), S fimbriae and F1C fimbriae (*sfa/foc*), afimbrial adhesin (*afa*), IbeA protein (*ibeA*), aerobactin (*iucC*), K1 capsule antigen (*neuC*), and type 1 fimbriae (*fimH*).

^b T_m, melting temperature.

by the unweighted pair group method with averages and are represented as a dendrogram.

Detection of VFs. PCR was used to detect eight genes encoding virulence determinants usually associated with the *E. coli* strains responsible for extra-intestinal infections: *neuC* (K1 capsule antigen), *hly* (alpha-hemolysin), *papC* (type P pili), *sfa/foc* (type S pili and type 1C fimbriae), *fimH* (type 1 pili), *afa* (afimbrial adhesin), *iucC* (aerobactin), and *ibeA* (IbeA protein) (2, 14, 18, 34, 35). Each VF gene was amplified with the primers described in Table 1 in a total volume of 25 µl containing 1× PCR buffer, 0.1 mM (each) deoxyribonucleoside triphosphate, 0.5 µM (each) primer, 0.5 U of *Taq* polymerase, and 25 ng of DNA. The reaction conditions were as follows: initial denaturation at 94°C for 3 min followed by 25 cycles of denaturation at 94°C for 1 min, annealing at the melting temperature of each primer (Table 1) for 1 min, and extension at 72°C for 1 min, followed by a final 10-min extension period at 72°C. The amplification products were separated by electrophoresis in a 1% agarose gel and visualized after ethidium bromide staining. A 100-bp DNA ladder (Life Technologies) was used in each gel as a molecular size marker. A negative control, consisting of the same mixture but with water instead of DNA, was included in each run. The results were considered to be positive if the amplification product was of the expected molecular size (Table 1). To confirm the positive results, at least five representative amplified products for each virulence gene were sequenced and compared with the sequences in the GenBank database. The products were sequenced by the dideoxyribonucleoside termination chain method (29), using an ABI Prism 377 DNA sequencer (Perkin-Elmer) and a ThermoSequenase dye terminator cycle-sequencing premix kit (Amersham Life Sciences, Cleveland, Ohio) according to the manufacturer's recommendations.

Statistical analysis. The distributions of the strains isolated from the various ecosystems and of the different pathogenic determinants among the genetic groups were analyzed. Differences in the distributions of the studied determinants were tested by the χ^2 test or by Fisher's exact test. A *P* value of < 0.05 was considered to indicate statistical significance.

RESULTS

Intraspecies distribution of 105 *E. coli* strains. (i) **ECOR grouping.** Among the 105 *E. coli* strains, 12 (11.4%) clustered in ECOR group A, 7 (6.7%) clustered in the B1 group, 71 (67.6%) clustered in the B2 group, and 15 (14.3%) clustered in the D group (Table 2).

(ii) **RAPD analysis.** The genetic structure of the population of 105 *E. coli* strains and the genetic links between strains were analyzed by RAPD analysis with two primers. Sixty-eight

TABLE 2. Distribution of 105 *E. coli* strains isolated from pregnant women and neonates in RAPD groups and four ECOR groups

ECOR group	No. (%) of strains			
	Total	RAPD group ^a		Other strains
		I	II	
A	12	1 (1.5)	9 (25)	2 (25)
B1	7	1 (1.5)	6 (17)	0
B2	71	54 (89)	12 (33)	5 (62)
D	15	5 (8)	9 (25)	1 (13)
Total	105	61 (58)	36 (34)	8 (8)

^a The clustering obtained by RAPD was significantly correlated with the distribution of strains in ECOR groups ($P < 0.00001$).

RAPD patterns were generated with the 188 primer, and 95 were generated with the 3388 primer. Primers 188 and 3388 generated a large number of DNA fragments per strain (3 to 11 and 2 to 15, respectively) including 18 and 15 polymorphic fragments, respectively. The molecular sizes of these fragments ranged from 520 to 3,200 bp for 188 and from 500 to 3,000 bp for 3388. By combining the RAPD patterns obtained with these two primers, 105 different RAPD types were observed for the 105 *E. coli* strains. The genetic relationships among these 105 RAPD types are shown in Fig. 1. The 105 strains diverged by up to 53%. At a level of 48% dissimilarity, three RAPD groups (I, II, and III) were determined. One strain did not cluster in any of the three groups (Fig. 1). Groups I and II accounted for a large part of the population, with 61 and 36 strains, respectively (Table 2).

A correlation was found between the intraspecies distributions of strains in the genogroups defined by ECOR PCR grouping and by RAPD analysis (Table 2). Indeed, 42% of the RAPD II group strains belonged to the ECOR A or B1 group, and 33% belonged to the ECOR B2 group, whereas 3% of the RAPD I group strains belonged to the ECOR A or B1 group and 89% belonged to the ECOR B2 group (i.e., the strains involved in extraintestinal infections). In contrast, the ECOR D group contained roughly equal numbers of strains belonging to the two major RAPD I and II groups.

Intraspecies distribution of *E. coli* strains according to their anatomical origins. The *E. coli* strains isolated from each anatomical site were not randomly distributed among the different ECOR groups with the exception of strains from the intestinal flora (Table 3). Strains from the ECOR B2 group seemed to be better suited to colonize the vagina. They represented 68% of the commensal *E. coli* strains isolated from this site. Moreover, the prevalence of ECOR B2 group strains was high at sites that are usually sterile: amniotic fluid (81%), blood culture (100%), and CSF (66%).

The distributions of the *E. coli* strains in the different RAPD groups according to their anatomical origins were more significantly different than their distributions in ECOR groups (Table 3). The strains from the natural ecosystem of *E. coli*, the intestinal flora, mostly belonged to the RAPD II group (79%), whereas only 17% of these strains belonged to the RAPD I group. *E. coli* strains from the vaginal flora were more equally distributed between the two major RAPD groups than among the ECOR groups. Fifty-two percent of the vaginal strains

belonged to the RAPD I group, and 40% belonged to the RAPD II group. Conversely, there was a strong correlation between the distributions of invasive strains in ECOR and RAPD groups. Eighty-four percent of the strains able to invade the uterine cavity (amniotic fluid) and the neonate (CSF and blood culture) belonged to the ECOR B2 group, and 79% belonged to the RAPD I group.

Distribution of the VF genes according to the anatomical origins of the strains. PCR was used to investigate the presence of *neuC*, *hly*, *fimH*, *iucC*, *papC*, *afa*, *sfa/foc*, and *ibeA* genes in all 105 strains. VF genes were detected in strains from all sites, but the prevalence in each population varied significantly (Table 4). Indeed, *E. coli* strains isolated from the intestinal flora contained few VF genes (median = 1.5 [range = 0 to 4]). *hly*, *sfa/foc*, *afa*, and *ibeA* were detected in only one of the 24 intestinal strains studied (4%), and *papC* was detected in only four of them (17%). The *iucC* (33%), *fimH* (58%), and *neuC* (42%) genes were more frequent in this population. In contrast, *E. coli* strains able to survive in the vagina possessed more VF genes (median = 4 [range = 0 to 6]). The prevalences of *neuC*, *fimH*, *iucC*, and *papC* were not significantly different in strains isolated from the vagina and in those isolated from the fecal flora. The vaginal isolates were more likely to possess the *hly* and *sfa/foc* genes than were the intestinal isolates (48 versus 4%; $P = 0.0006$). Little difference was observed between the VFs harbored by *E. coli* strains able to invade the amniotic fluid and the vaginal strains (median = 3 [range = 1 to 6]). Only the *fimH* gene was significantly more frequent in amniotic fluid strains (96%) than in vaginal strains (76%) and intestinal flora (58%) ($P = 0.03$).

Strains responsible for neonatal bacteremia contained a number of VF genes similar to that in strains isolated from the amniotic fluid (median = 4 [range = 3 to 6]), and the prevalences of each gene in these two populations were also comparable (Table 4).

The VFs harbored by the meningitis strains differed notably from those harbored by fecal strains. CSF strains were significantly more likely to possess the following genes than were the fecal strains: *neuC* (83 versus 42%; $P = 0.04$), *iucC* (92 versus 33%; $P = 0.003$), *sfa/foc* (33 versus 4%; $P = 0.05$), and *ibeA* (42 versus 4%; $P = 0.017$) (Table 4). The number of VF genes in CSF strains was comparable to that in blood culture strains (median = 4 [range = 1 to 5]), but CSF strains appeared to be less likely to harbor the *hly* (17 versus 65%; $P = 0.027$) and *fimH* (67 versus 100%; $P = 0.041$) genes than were blood culture strains.

Distribution of genes encoding pathogenic determinants among different genetic groups. Strains belonging to the phylogenetic ECOR B2 group and the intraspecies RAPD I group (Table 5), which contained most of the invasive strains (Table 3), contained the highest number of VF genes (median = 4 [range = 1 to 6] and median = 4 [range = 0 to 6], respectively). In addition, comparison of the prevalences of each VF gene between groups revealed particularities.

Significant differences in the prevalences of *neuC*, *hly*, *papC*, *sfa/foc*, and *ibeA* were observed among the four ECOR groups, and these five genes were most frequent in the ECOR B2 group (Table 5). Most of the strains (97 of 105) were distributed in one of the two major groups (RAPD I and II), and strains (56 of 105) isolated from usually sterile sites (amniotic

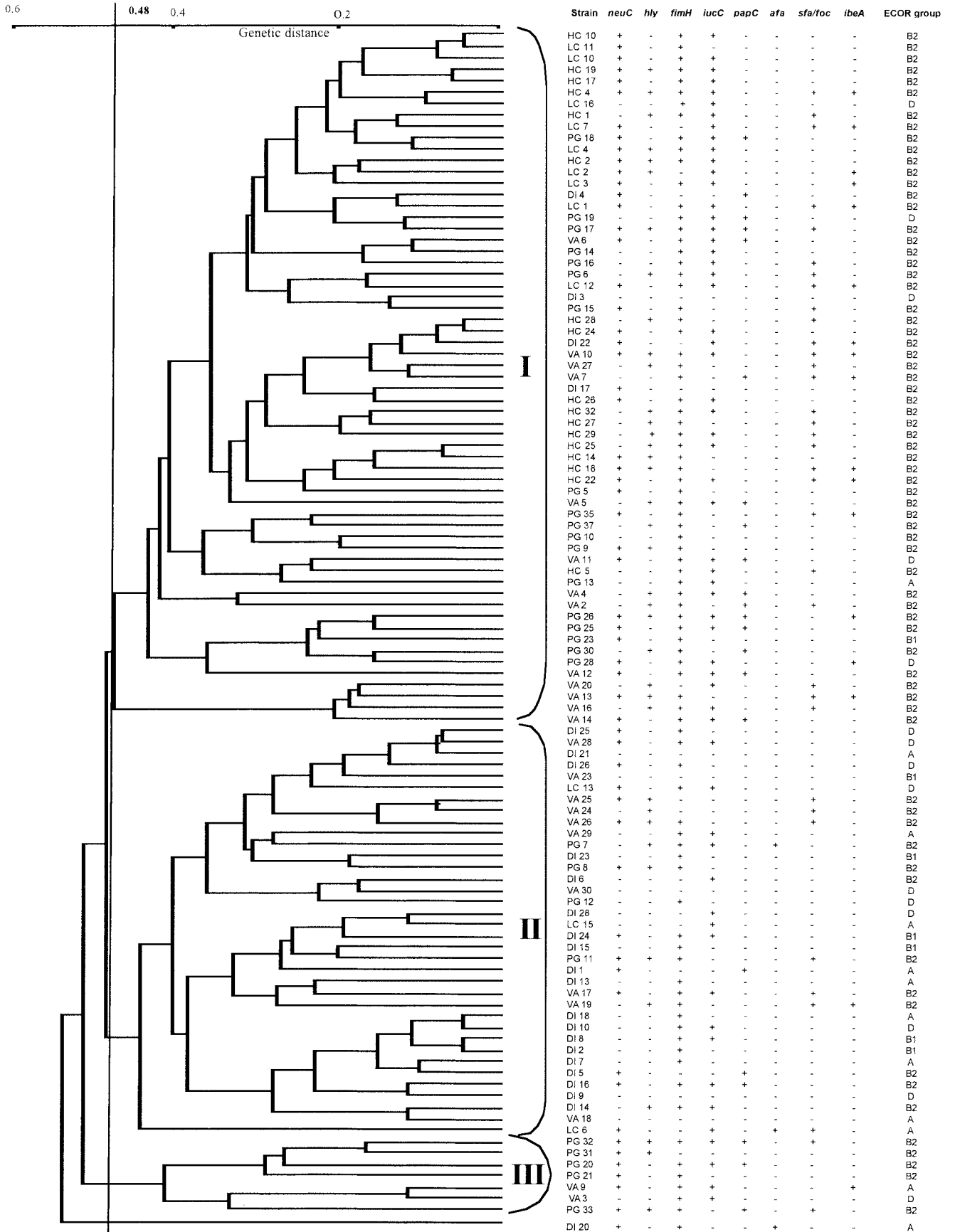


TABLE 3. Distribution of 105 *E. coli* strains isolated from pregnant women and neonates among ECOR groups and RAPD groups according to anatomical origin

Anatomical origin	No. (%) of strains ^a							
	Total	ECOR group ^b				RAPD groups ^c		Other strains
		A	B1	B2	D	I	II	
Intestinal flora	24	6 (25) ^c	5 (21)	7 (29)	6 (25)	4 (17)	19 (79)	1 (4)
Vaginal flora	25	3 (12)	1 (4)	17 (68)	4 (16)	13 (52)	10 (40)	2 (8)
Amniotic fluid	27	1 (4)	1 (4)	22 (81)	3 (11)	18 (67)	4 (15)	5 (18)
Blood culture	17	0	0	17 (100)	0	17 (100)	0	0
CSF	12	2 (17)	0	8 (66)	2 (17)	9 (75)	3 (25)	0
Total	105	12 (11)	7 (7)	71 (68)	15 (14)	61 (58)	36 (34)	8 (8)

^a Number (percent) of strains isolated from an anatomical origin that were classified in each ECOR and/or RAPD group or other group and total.

^b The distribution of strains isolated from each anatomical site in ECOR groups was not random ($P = 0.0024$).

^c The distribution of strains isolated from each anatomical site in RAPD groups was not random ($P = 0.00001$).

fluid, blood culture, and CSF) were clustered in the RAPD I group (Table 3). A comparison of the prevalences of individual VF genes in these two major RAPD groups showed that strains belonging to the RAPD I group were significantly more likely to harbor all of the VF genes except *afa* (Table 5).

DISCUSSION

E. coli and *S. agalactiae* are the major bacteria involved in neonatal septicemia and meningitis. These diseases can lead to death or serious neurological side effects (2, 4, 13). Most *E. coli* strains that cause neonatal meningitis and septicemia belong to the highly clonal ECOR B2 group and, to a lesser extent, to the ECOR D group. Conversely, most of the strains isolated from the fecal flora belong to the B1 and A groups (3, 9, 25, 26). The K1 capsular antigen, alpha-hemolysin, the iron uptake aerobactin system, adhesins (F1C fimbriae, P pili, and S pili), and the IbeA protein are considered to be the major VFs related to neonatal pathogenicity due to *E. coli* (2, 3, 13, 26).

It is generally hypothesized that *E. coli* strains responsible for amniotitis during pregnancy and at birth, and for early bacteremia and meningitis in neonates, are intestinal strains that have colonized the vaginal flora (17). Nevertheless, the exact genetic link between the commensal strains of these two anatomical sites and the invasive strains isolated from amniotic fluid, blood, and CSF cultures remains unclear. In this study, our goal was to identify this link to determine whether the vagina and/or the cervix favors colonization by strains that possess features different from those of fecal flora strains. If the selected strains are closely genetically related to those isolated from neonates, these human anatomical sites may then be considered as barriers that select for strains with a greater capacity to cause amniotitis and invasive diseases in neonates. In addition, a better knowledge of the characteristics

of genital and neonatal *E. coli* strains could help to identify particularities that could form the basis for developing tools able to identify strains that present a high risk in neonates. This would then allow physicians to propose screening strategies for prophylaxis during pregnancy.

We used two methods to characterize 105 strains of *E. coli*. The first, PCR ECOR grouping, is based on the search for two genes and an anonymous DNA fragment. This method was used to determine the phylogenetic positions of the *E. coli* strains in the four ECOR groups (7, 10). The second method, RAPD fingerprinting, explores the genome thoroughly. This is an interesting method for assessing the genetic structure of a microbial population and for phylogenetic analysis (8, 12, 13, 20, 28, 32). The results obtained with these two methods show two comparable genetic species structures (Table 2). However, the RAPD method seems to be better able to distinguish the population of *E. coli* strains isolated from intestinal flora from the population that presents a high risk for neonates (Table 3). PCR ECOR grouping confirmed that ECOR B2 strains present a high risk for neonates (66 to 100% of the neonatal invasive strains belonged to this group). However, according to this method 68% of the strains isolated from the vagina are also in this phylogenetic group (Table 3). In contrast, the RAPD method was better able to discriminate vaginal strains that may present a high risk of neonatal infection. Indeed, 75 to 100% of the strains isolated from neonatal invasive diseases were in the RAPD I group, whereas 52% of the vaginal flora strains were in this group (Table 3). Strains from the RAPD I group contained more VFs. Effectively, seven VF genes (*neuC*, *hly*, *fimH*, *iucC*, *papC*, *sfa/foc*, and *ibeA*) were significantly more frequent in the RAPD I group, whereas only five genes (*neuC*, *hly*, *papC*, *sfa/foc*, and *ibeA*) were significantly associated with the ECOR B2 group (Table 5). The distribution of VFs pro-

FIG. 1. Genetic relationships among 105 *E. coli* strains. Five distinct populations can be seen: 24 strains were isolated from the digestive tracts of asymptomatic pregnant women (DI), 25 were isolated from the vaginal flora of asymptomatic pregnant women (VA), 27 were isolated from amniotic fluid (PG), 17 were isolated from the blood of neonates without meningitis (HC), and 12 were isolated from the CSF of neonates (LC). The classification and the divergence of the isolates were calculated by the unweighted pair group method with averages using the RAPD data obtained with the 188 and 3388 primers. Three RAPD groups (I to III) were identified at 48% divergence. For each strain, the presence (+) or absence (–) of the eight genes encoding VFs (*neuC*, *hly*, *fimH*, *iucC*, *papC*, *afa*, *sfa/foc*, and *ibeA*) sought by PCR and the PCR ECOR grouping are noted.

TABLE 4. Distribution of genes encoding VFs^a

Anatomical origin	No. (%) of strains								
	Total	<i>neuC</i>	<i>hly</i>	<i>fimH</i>	<i>iucC</i>	<i>papC</i>	<i>afa</i>	<i>sfa/foc</i>	<i>ibeA</i>
Intestinal flora	24	10 (42)	1 (4)	14 (58)	8 (33)	4 (17)	1 (4)	1 (4)	1 (4)
Vaginal flora	25	11 (44)	12 (48)	19 (76)	14 (56)	8 (32)	0	12 (48)	5 (20)
Amniotic fluid	27	17 (63)	12 (44)	26 (96)	13 (48)	10 (37)	1 (4)	8 (30)	3 (11)
Blood culture	17	10 (59)	11 (65)	17 (100)	13 (76)	0	0	10 (59)	3 (18)
CSF	12	10 (83)	2 (17)	8 (67)	11 (92)	0	1 (8)	4 (33)	5 (42)
Total	105	58 (55)	38 (36)	84 (80)	59 (56)	22 (21)	3 (3)	35 (33)	17 (16)

^a K1 capsule antigen (*neuC*), alpha-hemolysin (*hly*), type 1 pili (*fimH*), aerobactin (*iucC*), type P pili (*papC*), afimbrial adhesin (*afa*), type S pili and F1C fimbriae (*sfa/foc*), and IbeA protein (*ibeA*) according to the anatomical origins of the strains. The data are expressed as numbers (percents) of strains carrying a given gene.

vides further evidence that members of the RAPD I group are more virulent than members of the ECOR B2 group (Table 5). Moreover, this result adds weight to the recent conclusions of Johnson et al. (13). ECOR B2 group *E. coli* strains isolated in newborn invasive diseases that cluster in the RAPD I group are probably a particularly virulent subpopulation. This result also suggests that F1 fimbriae and the aerobactin system play a peculiar role in the physiopathology of maternofetal infectious diseases caused by *E. coli*. In any case, if a subpopulation of strains that is particularly virulent in neonates exists in the ECOR B2 group, the RAPD grouping method is suitable for identifying this subgroup. However, a high level of standardization and internal controls are necessary to ensure good reproducibility.

It is generally thought that *E. coli* strains involved in neonatal infections originate from the vagina, which was in turn colonized from a rectal source (17). If this is true, our data indicate that the physiological conditions of the different ecosystems crossed impose a strong selection on the bacteria and that this pressure leads to progressive changes in the dominant *E. coli* subpopulation. Indeed, a large part of the RAPD II group consisted of *E. coli* strains from the intestinal flora, whereas most of the strains isolated in maternofetal and neonatal diseases (amniotitis, bacteremia, and meningitis) belonged to the RAPD I group. This population change was first imposed by the vaginal ecosystem. The vaginal *E. coli* strains

harbored a lot more VFs than did the *E. coli* strains isolated from the fecal flora (median = 4 versus 1.5) (Table 4). Thus, the vagina acts as a partial filter that favors colonization by strains that present a high risk of infection for newborns, i.e., members of the RAPD I group, which contained 52% of the vaginal strains tested. The second anatomical site that selects for *E. coli* strains that present a high risk for neonates is the endocervix and/or the amniotic fluid: 67% of the *E. coli* strains that invaded the amniotic cavity through the endocervix belonged to RAPD group I (Table 3).

The ability of some *E. coli* strains to adapt to the vaginal conditions and the amniotic fluid and to invade the blood or the CSF of neonates correlates with an enrichment of VFs (Tables 4 and 5). Two genes, *iucC* and *ibeA*, were found to be especially associated with CSF strains and may play major roles in the physiopathology of meningitis. Aerobactin, an iron transport system (23), allows bacterial growth in the host, and the IbeA protein, which interacts with a protein on the surfaces of brain microvascular endothelial cells, may favor invasion of the central nervous system (11, 15). Two other genes, *hly* and *sfa/foc*, were found to be associated with strains isolated from blood culture and thus may be of importance in septicemia. Alpha-hemolysin can lyse erythrocytes, leading to the release of intracellular iron (16, 19), and the type S pili-type1C fimbriae promote adhesion to microvascular endothelial cells (24, 27).

TABLE 5. Distribution of genes encoding VFs among RAPD and ECOR groups^a

VF gene	No. of genes in ECOR group ^b :				<i>P</i> ^c	No. of genes in RAPD group ^b :		<i>P</i> ^d	No. of genes in other strains ^b (n = 8)
	A (n = 12)	B1 (n = 7)	B2 (n = 71)	D (n = 15)		I (n = 61)	II (n = 36)		
<i>neuC</i>	4 (33)	2 (29)	46 (65)	6 (40)	0.0045	37 (61)	14 (39)	0.04	7 (87)
<i>hly</i>	0	0	38 (53)	0	<0.0001	27 (44)	8 (22)	0.03	3 (37)
<i>fimH</i>	7 (58)	6 (86)	60 (84)	11 (73)	NS ^e	54 (88)	23 (64)	0.004	7 (87)
<i>iucC</i>	5 (42)	2 (29)	43 (61)	9 (60)	NS	41 (67)	14 (39)	0.007	4 (50)
<i>papC</i>	1 (8)	0	19 (27)	2 (13)	0.033	16 (26)	3 (8)	0.03	3 (37)
<i>afa</i>	2 (17)	0	1 (1)	0	NS	0	2 (6)	NS	1 (12)
<i>sfa/foc</i>	1 (8)	0	34 (48)	0	<0.0001	26 (43)	7 (19)	0.018	2 (25)
<i>ibeA</i>	1 (8)	0	15 (21)	1 (7)	0.046	15 (25)	1 (3)	0.005	1 (12)
Median (range)	2 (0-4)	1 (0-3)	4 (1-6)	2 (0-4)		4 (0-6)	2 (0-4)		3.5 (2-6)

^a K1 capsule antigen (*neuC*), alpha-hemolysin (*hly*), type 1 pili (*fimH*), aerobactin (*iucC*), type P pili (*papC*), afimbrial adhesin (*afa*), type S pili and F1C fimbriae (*sfa/foc*), and IbeA protein (*ibeA*).

^b Number (percent) of strains positive for the virulence gene as demonstrated by a PCR assay.

^c ECOR group B2 versus all other ECOR groups.

^d RAPD group I versus RAPD group II.

^e NS, not significant.

There are some similarities between vaginal colonization and neonatal infection with *S. agalactiae* and *E. coli*. Thus, programs to screen for vaginal colonization by *E. coli* may be useful in pregnant women. Nevertheless, our data indicate that only about one-half of the vaginal *E. coli* strains belong to the highly virulent (RAPD I) group. Therefore, tools able to identify *E. coli* strains that present a high risk for neonates need to be developed to avoid unjustified antibiotic therapy. Given that RAPD grouping was better able to distinguish the highly virulent subpopulation of strains responsible for neonatal invasive diseases than was ECOR grouping, these tools should be tested for the ability to recognize RAPD I group strains and not only ECOR B2 group strains. However, the lack of inter-laboratory reproducibility associated with the RAPD method means that further techniques should be developed to screen for this particularly virulent subpopulation.

In conclusion, if *E. coli* strains responsible for invasive neonatal infections come from the intestinal flora, our data suggest that the vagina in pregnant women and the amniotic fluid are two barriers that favor the selection of a population of *E. coli* strains that present a high risk of infection for neonates. This population appears to constitute a highly virulent subpopulation within the ECOR B2 phylogenetic group.

ACKNOWLEDGMENTS

The *E. coli* strains isolated from blood culture and CSF were provided by the Collège de Bactériologie, Virologie et Hygiène des Hôpitaux. We thank Stéphanie Gouriou for technical assistance.

REFERENCES

- Amstey, M. S., E. Lewin, and J. Colaice. 1980. Vaginal colonization with invasive *Escherichia coli* during pregnancy. *Am. J. Obstet. Gynecol.* **137**:534–535.
- Bingen, E., S. Bonacorsi, N. Brahimi, E. Denamur, and J. Elion. 1997. Virulence patterns of *Escherichia coli* K1 strains associated with neonatal meningitis. *J. Clin. Microbiol.* **35**:2981–2982.
- Bingen, E., B. Picard, N. Brahimi, S. Mathy, P. Desjardin, J. Elion, and E. Denamur. 1998. Phylogenetic analysis of *Escherichia coli* strains causing neonatal meningitis suggests horizontal gene transfer from a predominant pool of highly virulent B2 group strains. *J. Infect. Dis.* **177**:642–650.
- Bradley, J. S. 1985. Neonatal infections. *Pediatr. Infect. Dis.* **4**:315–320.
- Cave, H., E. Bingen, J. Elion, and E. Denamur. 1994. Differentiation of *Escherichia coli* strains using randomly amplified polymorphic DNA analysis. *Res. Microbiol.* **145**:141–150.
- Chatellier, S., C. Ramanantsoa, P. Harriau, K. Rolland, A. Roseneau, and R. Quentin. 1997. Characterization of *Streptococcus agalactiae* strains by randomly amplified polymorphic DNA analysis. *J. Clin. Microbiol.* **35**:2573–2579.
- Clermont, O., S. Bonacorsi, and E. Bingen. 2000. Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl. Environ. Microbiol.* **66**:4555–4558.
- Desjardin, P., B. Picard, B. Kaltenböck, J. Elion, and E. Denamur. 1995. Sex in *Escherichia coli* does not disrupt the clonal structure of the population: evidence from random amplified polymorphic DNA and restriction-fragment-length polymorphism. *J. Mol. Evol.* **41**:440–448.
- Duriez, P., O. Clermont, S. Bonacorsi, E. Bingen, A. Chaventre, J. Elion, B. Picard, and E. Denamur. 2001. Commensal *Escherichia coli* isolates are phylogenetically distributed among geographically distinct human populations. *Microbiology* **147**:1671–1676.
- Herzer, P. J., S. Inouye, M. Inouye, and T. S. Whittam. 1990. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *J. Bacteriol.* **172**:6175–6181.
- Huang, S. H., C. Wass, O. Fu, N. V. Prasadarao, M. Stins, and K. S. Kim. 1995. *Escherichia coli* invasion of brain microvascular endothelial cells in vitro and in vivo: molecular cloning and characterization of invasion gene *ibeA*. *Infect. Immun.* **63**:4470–4475.
- 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.
- Johnson, J. R., E. Oswald, T. T. O'Bryan, M. A. Kuskowski, and L. Spanjaar. 2002. Phylogenetic distribution of virulence-associated genes among *Escherichia coli* isolates associated with neonatal bacterial meningitis in The Netherlands. *J. Infect. Dis.* **185**:774–784.
- 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.
- Kim, K. S. 2001. *Escherichia coli* translocation at the blood-brain barrier. *Infect. Immun.* **69**:5217–5222.
- Korhonen, T. K., M. V. Valtonen, J. Parkkinen, V. Vaisanen-Rhen, J. Finne, F. Orskof, I. Orskof, S. B. Svenson, and P. H. Makela. 1985. Serotypes, hemolysin production, and receptor recognition of *Escherichia coli* strains associated with neonatal sepsis and meningitis. *Infect. Immun.* **48**:486–491.
- Krohn, M. A., S. S. Thwin, L. K. Rabe, Z. Brown, and S. L. Hillier. 1997. Vaginal colonization by *Escherichia coli* as a risk factor for very low birth weight delivery and other perinatal complications. *J. Infect. Dis.* **175**:3606–3610.
- Le Bouguenec, C., M. Archambaud, and A. Labigne. 1992. Rapid and specific detection of the *pap*, *afa*, and *sfa* adhesin-encoding operon in uropathogenic *Escherichia coli* strains by polymerase chain reaction. *J. Clin. Microbiol.* **30**:1189–1193.
- Maslow, J. N., T. S. Whittam, C. F. Gilks, R. A. Wilson, M. E. Mulligan, K. S. Adams, and R. D. Arbeit. 1995. Clonal relationships among bloodstream isolates of *Escherichia coli*. *Infect. Immun.* **63**:2409–2417.
- Mereghetti, L., P. Lanotte, V. Savoye-Marczuk, N. Marquet-Van Der Mee, A. Audurier, and R. Quentin. 2002. Combined ribotyping and random multiplex DNA analysis to probe the population structure of *Listeria monocytogenes*. *Appl. Environ. Microbiol.* **68**:2849–2857.
- Ochman, H., and K. Selander. 1984. Evidence for clonal population structure in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **81**:198–201.
- Ochman, H., and K. Selander. 1984. Standard reference strains of *Escherichia coli* from natural populations. *J. Bacteriol.* **157**:690–693.
- Opal, S. O., A. S. Cross, P. Gemski, and L. W. Lyhte. 1990. Aerobactin and alpha hemolysin as virulence determinants in *Escherichia coli* isolated from human blood, urine and stool. *J. Infect. Dis.* **161**:794–796.
- Parkkinen, J., T. K. Korhonen, A. Pere, J. Hacker, and S. Soinila. 1988. Binding sites in the rat brain for *Escherichia coli* S fimbriae associated with neonatal meningitis. *J. Clin. Invest.* **81**:860–865.
- Picard, B., N. Picard-Pasquier, R. Krishnamoorthy, and P. Goullet. 1991. Characterization of highly virulent *Escherichia coli* strains by ribosomal DNA restriction fragment length polymorphism. *FEMS Microbiol. Lett.* **82**:183–188.
- Picard, B., J. Sevali Garcia, S. Gouriou, P. Duriez, N. Brahimi, E. Bingen, J. Elion, and E. Denamur. 1999. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect. Immun.* **67**:546–553.
- Prasadarao, N. V., C. A. Wass, and K. S. Kim. 1997. Identification and characterization of S fimbria-binding sialoglycoproteins on brain microvascular endothelial cells. *Infect. Immun.* **65**:2852–2860.
- Ruimy, R., E. Genauzeau, C. Barnabé, A. Beaulieu, M. Tybayrenc, and A. Andreumont. 2001. Genetic diversity of *Pseudomonas aeruginosa* strains isolated from ventilated patients with nosocomial pneumonia, cancer patients with bacteremia, and environmental water. *Infect. Immun.* **69**:584–588.
- Sanger, F. 1975. Determination of nucleotide sequences in DNA. *Science* **214**:1205–1210.
- Sarff, L. D., and G. H. McCracken. 1975. Epidemiology of *Escherichia coli* K1 in healthy and diseased newborns. *Lancet* **i**:1099–1104.
- Selander, R. K., T. K. Korhonen, V. Vaisanen-Rhen, P. H. Williams, P. E. Pattison, and D. A. Caugant. 1986. Genetic relationships and clonal structure of strains of *Escherichia coli* causing neonatal septicemia and meningitis. *Infect. Immun.* **52**:213–222.
- Wang, G., T. S. Whittam, C. M. Berg, and D. E. Berg. 1993. RAPD (arbitrary primer) PCR is more sensitive than multilocus enzyme electrophoresis for distinguishing related bacterial strains. *Nucleic Acids Res.* **21**:5930–5933.
- Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18**:6531–6535.
- Yamamoto, S., A. Terai, K. Yuri, H. Kurazono, Y. Takeda, and O. Yoshida. 1995. Detection of urovirulence factors in *Escherichia coli* by multiplex polymerase chain reaction. *FEMS Immunol. Med. Microbiol.* **12**:85–90.
- Zapata, G., J. M. Crowley, and W. F. Vann. 1992. Sequence and expression of the *Escherichia coli* *neuC* gene product. *J. Bacteriol.* **174**:315–319.