INFECTION AND IMMUNITY, July 2000, p. 3983–3989 Vol. 68, No. 7
0019-9567/00/$04.00+0


# Prevalence of Virulence Genes and Clonality in *Escherichia coli* Strains That Cause Bacteremia in Cancer Patients

FARIDA HILALI,[1,2] RAYMOND RUIMY,[1] PATRICK SAULNIER,[2] CHRISTIAN BARNABÉ,[3] CHANTAL LEBOUGUÉNEC,[4] MICHEL TIBAYRENC,[3] AND ANTOINE ANDREMONT[1]*

*EMI INSERM 9933, AP-HP Hôpital Bichat Claude-Bernard,[1] and Unité de Pathogénie Bactérienne des Muqueuses, Institut Pasteur,[4] Paris, Laboratoire de Microbiologie Médicale, Institut Gustave Roussy, Villejuif,[2] and Centre d'Etudes sur le Polymorphisme des Microorganismes (CEPM), UMR CNRS/IRD 9926, IRD, Montpellier,[3] France*

Received 20 April 1999/Returned for modification 30 June 1999/Accepted 31 March 2000

**Phenotypic analysis of *Escherichia coli* strains causing bacteremia in cancer patients suggests that they possess specific virulence properties. To investigate this hypothesis, we compared the frequency of the virulence-related genes *cnf1*, *cnf2*, *papC*, *hlyC*, and *iut* in 155 *E. coli* strains isolated from hospitalized cancer patients with epidemiologically unrelated cases of bacteremia to their frequency in 70 *E. coli* strains isolated from the feces of healthy unrelated volunteers. Of the blood isolates, 24, 37, and 26% were positive for *cnf1*, *papC*, and *hlyC*, respectively, versus only 6, 17, and 6% of the fecal isolates ($P < 0.05$ in all instances). By contrast, 47% of both isolates carried the *iut* gene. The patients' clinical characteristics did not significantly influence these frequencies. The presence on various pathogenicity islands (PAIs) of a combination of the *cnf1*, *papC*, and *hlyC* genes on the chromosome was strongly suggested by Southern blotting of pulsed-field gel electrophoresis (PFGE) patterns with specific DNA probes. The phylogenetic relatedness among 60 strains carrying three, two, one, or no virulence genes and 6 ECOR strains included as references was determined by neighbor joining, the unweighted pair-group method with arithmetic mean, and Wagner analysis of the randomly amplified polymorphic DNA (RAPD) patterns generated by 11 primers. Identification of a major cluster including 96.4% of the strains carrying the *cnf1*, *papC*, and *hlyC* genes and ECOR subgroup B2 strains suggested that the virulent *E. coli* strains causing bacteremia in cancer patients are closely related to ECOR B2 strains. The presence in the *E. coli* population surveyed of a strong linkage disequilibrium, and especially of a highly significant correlation between PFGE and RAPD genetic distances, confirms that clonal propagation has a major impact on the *E. coli* population structure. Nevertheless, low bootstrap values in the phylogenetic tree suggested that frequent genetic exchange inhibits the individualization of discrete genetic lineages, which are stable on an evolutionary scale.**

Patients with cancer are highly susceptible to bacterial infections, particularly during the periods of severe neutropenia that follow anticancer chemotherapy. Bloodstream infections caused by gram-negative bacteria are among the most frequent of these infections and are associated with a high mortality rate (1). Most of the gram-negative bacteria that cause these bloodstream infections originate in the intestinal flora (55) and reach the blood by a mechanism called bacterial translocation (for a review, see reference 4). In experimental models, bacterial translocation is enhanced by intestinal colonization and overgrowth of the translocating bacteria, by immunosuppression of the host, and by alterations of the intestinal mucosa (for a review, see reference 4), all of which may occur in cancer patients.

*Escherichia coli* is the most common gram-negative species isolated from cancer patients with bacteremia (1, 55). Phenotypic studies performed on a series of strains isolated from the blood of patients with various underlying conditions suggest that the *E. coli* strains isolated from such patients express traits associated with virulence more frequently than do the strains isolated from the intestinal flora of healthy volunteers. These traits include O serogroup (13), alpha-hemolysin (2, 7, 13, 43), type B carboxylesterase (13, 21), cytotoxic necrotizing factor (CNF) (13), aerobactin (2, 13, 43), and colicin (13). This was confirmed by genotypic studies. For instance, *hly*, the operon encoding alpha-hemolysin, *aero*, the operon which encodes aerobactin, *cnf1*, which encodes CNF1, and *pap*, the operon which encodes pyelonephritis adhesin pili, were found preferentially in *E. coli* strains isolated from patients with urinary tract infections (8, 43, 64); *pap*, *hly*, *aero*, and *afa*, the operon encoding afimbrial adhesins, were found in patients with neonatal meningitis (5); and *pap*, *hly*, *aero*, and *sfa*, the operon which encode S fimbria adhesins, were found in patients with bacteremia (35, 43). *cnf2*, the gene encoding CNF2, has been found only in strains isolated from animals (45).

Virulence genes are often grouped in a single segment of the bacterial chromosome, constituting a distinct functional genetic unit called a pathogenicity island (PAI) (10, 24). PAIs have now been isolated from many bacterial species (for a recent review, see reference 25). They are often larger than 30 kb, have a G+C content different from that of the host organism, are often inserted into tRNA genes, may be flanked by insertion sequence elements, may be lost by deletion at high frequency, and, typically, are present more often in pathogenic than in less pathogenic strains within the same species (25). *E. coli* PAIs have been studied chiefly in uropathogenic and enteropathogenic strains. Several PAIs, all carrying *hly* genes and numbered I, II, IV, and V, have been described in uropathogenic strains (10, 54). In addition to *hly*, PAI II contains *prf*, encoding the P-related fimbriae (10), PAI IV contains *pap* (54), and PAI V contains both *prs* encoding a P-related sequence and *cnf1* (10). PAI III, a newly renamed locus of enterocyte effacement, has been identified in enteropathogenic

* Corresponding author. Mailing address: Laboratoire de Bactériologie, Groupe Hospitalier Bichat-Claude Bernard, 47, rue Henri Huchard, 75877 Paris Cedex 18, France. Phone: 33 (0)1 40 25 85 00. Fax: 33 (0)1 40 25 85 81. E-mail: antoine.andremont@bch.ap-hop-paris.fr.

*E. coli* strains. It carries a specific set of virulence genes comprising *sepA* through *sepI*, encoding a type III secretion system protein, *espA* and *espB*, encoding secreted proteins involved in signal transduction, and *eaeA*, encoding intimin (37, 38). So far, the presence of PAIs has not been reported in *E. coli* strains causing bloodstream infections.

The population structure of *E. coli*, as represented by the ECOR collection (41), is thought to be clonal, since it comprises five major clonal groups called A, B1, B2, D, and E (26). Most *E. coli* blood isolates belong to the B2 ECOR group (48). This is also true for *E. coli* strains isolated from patients with neonatal meningitis (5). Recently, it was shown that the B2 and D ECOR strains carried virulence genes, including *hly*, *pap*, *sfa*, and *kps*, a cluster of genes encoding proteins involved in export of type II capsular polysaccharides, more often than did strains from the other ECOR groups (11). It was also reported that a set of virulence genes comprising *pap*, *hly*, and *sfa* was preferentially distributed in a major cluster of strains causing bloodstream infections in patients with various underlying conditions (35). All these studies suggest that a cluster of *E. coli* strains acquired virulence genes by horizontal transfer, thus defining a highly virulent group. The link between the B2 phylogenetic group and the virulence in *E. coli* strains causing extraintestinal infection was recently confirmed in studies of experimentally infected mice (47).

In the present study, we attempted to determine whether a clustered distribution of virulence genes can be found in *E. coli* strains causing bacteremia in cancer patients, and we used new approaches to explore further the general problem of the population structure of *E. coli* and its mode of evolution.

## MATERIALS AND METHODS

**Bacterial strains and patient characteristics.** We studied 155 strains of *E. coli* isolated between April 1992 and December 1993 from consecutive patients with bacteremia who were hospitalized at Institut Gustave-Roussy, a 420-bed French cancer referral center. The major underlying diseases were leukemia in 8 patients (5.2%), lymphoma in 17 (11%), solid tumors in 114 (73.5%) and unknown in 16 (10.3%). Neutropenia was defined as a leukocyte count lower than 500/mm$^3$ at the onset of bacteremia, and a urinary source of bacteremia was suspected when more than $10^5$ CFU of *E. coli* per ml was isolated from urine samples. We also studied 70 *E. coli* strains isolated from the feces of healthy volunteers living in the community. For each fecal sample, one colony was chosen at random and it was assumed that the isolates of *E. coli* so obtained were representative of the bacterial population present in the feces of the volunteers at that time.

*E. coli* reference strains isolated from humans or animals, with or without known virulence genes, were used as positive and negative controls for the PCR studies. Six *E. coli* strains from the ECOR collection (41) were used as reference strains for the phylogenetic study, and three unrelated strains of enterobacteria were taken as the outgroup (Table 1).

**Bacterial cultures and DNA extraction.** *E. coli* strains were grown on Luria-Bertani agar or in Luria-Bertani broth at 37°C overnight. Genomic DNA was extracted by rapid lysis, as described previously (50), and used as the target for PCR assays.

**PCR materials and methods.** Primers were used to amplify intragenic fragments of the genes *cnf1* (15), *cnf2* (46), *papC* (32), *hlyC* (17), *iut* (representative of the aerobactin operon) (13b), and the tRNA *pheR*, which encodes phenylalanine tRNA and which is often the site of insertion from PAI V. Details of primer sequences and predicted sizes of the amplified products are given in Table 2. PCR amplification of bacterial DNA extracts was performed in 25-μl volumes containing 5 μl of cell lysate for detecting *cnf1* and *cnf2* (51) and *papC* (32). For detection of *hlyC* and *iut*, PCRs were performed using 15 mM Tris-HCl (pH 8.3), 40 mM KCl, 200 μM each deoxynucleoside triphosphate (Boehringer Mannheim, Meylan, France), 1 μM each primer, and 1 U of Ampli-Taq DNA polymerase (Perkin-Elmer) per reaction. The concentration of $MgCl_2$ optimized for each PCR was 4 mM for *hlyC* and 1 mM for *iut*. PCR amplification consisted of one cycle at 94°C for 5 min, 28 cycles of denaturation at 94°C for 45 s, annealing at 69°C for 40 s, and extension at 72°C for 1 min, and one cycle of 72°C for 5 min in a Thermocycler 9600 (Perkin-Elmer/Cetus). For the detection of tRNA *pheR*, the PCR mixtures differed from those of *hlyC* and *iut* in that Ampli-Taq DNA was replaced by DNA *Taq* polymerase (Boehringer Mannheim). PCRs were performed in a final volume of 50 μl containing 5 μl of cell lysate. The concentration of $MgCl_2$ was 1 mM. PCR amplification consisted of one cycle at 94°C for 5 min, 25 cycles of denaturation at 94°C for 30 s, annealing at 52°C for 45 s, and extension 72°C for 1 min, and one cycle of 72°C for 10 min.

TABLE 1. Characteristics of the *E. coli* reference strains isolated from humans and animals

| Strain | Phenotype | ECOR subgroup | Host | Reference |
|---|---|---|---|---|
| J96 | Cnf1, Pap, Hly | ND[a] | Human | 29 |
| 1520 | Cnf1 | ND | Dog | 45 |
| BM4.1 | Cnf1 | ND | Calf | 45 |
| BM2.1 | Cnf1 | ND | ND | 45 |
| JL21+ | Cnf2 | ND | Calf | 45 |
| S5 | Cnf2, Aero | ND | Lamb | 45 |
| BM2.10 | Cnf1, Aero | ND | Calf | 45 |
| HB101 | No known virulence factor | ND | Human | 45 |
| JL21− | JL21 without Cnf2 | ND | Calf | 45 |
| ECOR4 | ND | A | Human | 42 |
| ECOR25 | ND | A | Dog | 42 |
| ECOR45 | ND | B1 | Pig | 42 |
| ECOR48 | Pap, Hly | D | Human | 42 |
| ECOR52 | Cnf1, Pap, Hly | B2 | Orangutan | 42 |
| ECOR59 | ND | B2 | Human | 42 |

[a] ND, not determined.

Aliquots (10 μl) of the final reaction mixtures underwent standard gel electrophoresis in 2% agarose. Amplified DNA fragments of specific sizes were detected by UV fluorescence after being stained with ethidium bromide.

**PFGE typing.** Epidemiological relationships between strains were assessed by studying the pulsed-field gel electrophoresis (PFGE) patterns of genomic DNA after restriction by *Xba*I, *Sfi*I, and *Not*I (Boehringer, Mannheim, Germany), as described previously (22, 23). These enzymes do not have restriction sites within the sequences of the five virulence genes under study. An *Xba*I digest of *Salmonella enterica* serovar Indiana (T461) DNA was used as a reference pattern for strain analysis and band size determination. PFGE patterns were compared using Gel-Compar software (Applied Maths, Kortrijk, Belgium). Strains with more than seven distinct restriction fragments were considered different, as described previously (56).

**Southern blot hybridization.** PFGE digests were transferred to nylon membranes (Amersham) and hybridized with digoxigenin-labeled probes which were generated by the specific primers of the *cnf1*, *papC*, and *hlyC* genes as recommended by the manufacturer (Boehringer Mannheim).

**Relationships among strains from our collection.** A subset of 60 strains from our collection were selected so as to include representative combinations of the virulence genes under study. We chose 19, 13, 8, and 20 bacteremic strains among those that carried three, two, one, and no virulence genes, respectively. DNA, extracted as described previously (33), was subjected to 11 different randomly amplified polymorphic DNA (RAPD) assays using the following commercially available 10-mer primers: A1, A2, A3, A7, A9, A10, A13, A16, A17, A19, and

TABLE 2. Oligonucleotide primers used to amplify virulence-associated genes

| Target gene | Oligonucleotide sequence (5′→3′) | Predicted size of amplified product (bp) | Reference |
|---|---|---|---|
| *cnf1* | CATGCTTCTTCCTCAGTAGC<br>GTCTAAAAGGGGGGCAGCCA | 255 | 51 |
| *cnf2* | CTTCTTCAGTTGTTCCTCCG<br>TAAGGCGGGCACAAACCAGA | 583 | 51 |
| *hlyC* | GCCAGAACAGTGCTTATCCT<br>CCAGTTCCCCATTACACAGA | 1,576 | This work |
| *iut* | AGGTGACGGCTTTCTCTTCG<br>TCTTGTTGTTCAGCCCTCCG | 189 | This work |
| *papC* | GACGGCTGTACTGCAGGGTGTGGCG<br>ATATCCTTTCTGCAGGGATGCAATA | 328 | 32 |
| tRNA *pheR* | AGCCGCGCTTTGGTACAGTAGC<br>CAGATGGAACTGGTGCTGGAAGG | 1,168 | This work |

TABLE 3. Prevalence of virulence genes in *E. coli* strains isolated from the blood of cancer patients and the feces of healthy volunteers

| Target gene | No. (%) of positive strains in: | | *P* value[a] |
|---|---|---|---|
| | Blood ($n$ = 155) | Feces ($n$ = 70) | |
| *cnfI* | 37 (24) | 4 (6) | <0.05 |
| *papC* | 58 (37) | 12 (17) | <0.05 |
| *hlyC* | 40 (26) | 4 (6) | <0.05 |
| *iut* | 73 (47) | 33 (47) | NS[b] |

[a] Chi-square test.
[b] NS, not significant.

A20 (kit A from Operons Technologies, Alameda, Calif.). Preliminary results (not shown) showed that these primers generated distinct reproducible RAPD patterns. RAPD assays were performed in duplicate in 61-μl reaction volumes containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 100 μM each deoxynucleoside triphosphate (Boehringer), 200 nM primer, 0.9 U of DNA *Taq* polymerase (Boehringer), and 10 ng of bacterial DNA. A PTC 100 DNA Engine machine (MJ Research, Watertown, Mass.) was used for amplification, which comprised 45 cycles, each consisting of 1 min at 94°C, 1 min at 36°C, and 2 min at 72°C. Amplified fragments underwent standard gel electrophoresis in 1.6% agarose and were detected by UV fluorescence after staining with ethidium bromide. Only the major bands of the resulting patterns were taken into account in the analysis (63).

We used two previously described genetic population tests, namely, the f and g tests (62), both based on the assessment of linkage disequilibrium (i.e., the nonrandom association of genotypes occurring at different loci) by Monte Carlo simulations with $10^4$ iterations. As a null hypothesis, both tests take panmixia, a situation in which genetic exchange occurs randomly. Means to avoid biases due to geographical and/or temporal separation have been proposed (57, 61). The f test only evaluates the probability of observing, by chance, a linkage disequilibrium as high as or higher than the one actually observed in the sample, whereas the g test also evaluates the correlation between the genetic distances calculated with two independent sets of genetic markers (here PFGE and RAPD patterns). In a panmixia, these two sets of markers recombine independently and therefore do not correlate. The correlation is evaluated by the nonparametric Mantel test (34), which does not require an assumed degree of freedom in sample.

The unweighted pair group method with arithmetic mean (53), neighbor-joining, and Wagner phylogenetic methods derived from the PHYLIP software package (19) were used to assess the phylogenetic relationships between the strains. *S. enterica* serovar Indiana and *Citrobacter freundii* were chosen as outgroups to root trees because they are closely related to the genus *Escherichia*. Only branch topologies found similar by all the above three methods were considered significant, because agreement between the results of several phylogenetic methods is a strong indication that phylogeny has been correctly determined (28, 31). The robustness of the branches was estimated by the bootstrap method (18).

## RESULTS

**Prevalence of virulence genes.** The *cnf1*, *papC*, and *hlyC* genes, but not the *iut* gene, were detected in a significantly larger number of *E. coli* blood isolates from cancer patients than in fecal isolates from healthy volunteers (Table 3). The *cnf2* gene was not detected in any human strains, as reported previously (45). As shown in Table 4, four different combinations of the genes under study were identified. Only the association of *cnf1*, *papC*, and *hlyC* occurred in significantly more blood than fecal isolates (19 and 4.2%, respectively; $P < 0.05$). Conversely, the absence of all virulence genes as detected here was noted in more fecal than blood isolates (81 and 57%, respectively; $P < 0.05$) (data not shown).

**Clinical data analysis.** We did not find any significant differences between the number of virulence factors in strains recovered from patients with or without the clinical characteristics studied, which included the number of circulating polymorphonuclear leukocytes at the onset of bacteremia, the underlying disease, and the association of bacteremia with a positive urine culture (data not shown).

**Epidemiological relationships between *E. coli* strains.** Digestion of *E. coli* DNA with *Xba*I, *Sfi*I, and *Not*I produced 16 to 20 well-resolved fragments ranging from 20 to 700 kb (data not shown). PFGE patterns were highly heterogeneous among the different strains. PFGE resolved at least seven distinct fragments when the patterns were compared pairwise. Thus, on the basis of widely accepted interpretative criteria (56), we concluded that the strains were epidemiologically unrelated.

**Physical location of virulence genes.** The results of Southern blotting of PFGE patterns with the *hlyC*, *cnf1*, and *papC* probes suggested that these three genes were physically linked in 22 of the 32 strains (69%) carrying the three virulence genes (Table 4). The 32 strains were found in either blood or fecal isolates. In 18 of 22 strains (82%), the three probes hybridized on the same restriction fragment with all three restriction enzymes. For the 10 strains carrying two virulence genes (Table 4), the probes hybridized with all the restriction enzymes in 7 of the 10 strains (70%). The sizes of the hybridization bands in the strains carrying three and two virulence genes were smaller than 190 kb in 10 of 22 strains (45%) and 4 of 10 strains (40%), respectively. This cutoff limit was chosen because the largest previously described PAI in *E. coli*, PAI II, is 190 kb long (9). Among the 22 strains carrying three virulences genes, the tRNA *pheR* was conserved in size in only 2 strains (10%). In comparison, 5 of 12 strains (42%) carrying no studied virulence genes had tRNA *pheR*.

**Population genetic and phylogenetic analysis.** The RAPD patterns generated by each primer were polymorphic (results not shown), and there was a highly significant correlation between RAPD and PFGE genetic distances ($r = 0.33$, $P < 10^{-4}$; Mantel nonparametric test). The linkage disequilibrium f and g tests were significant for both the overall sample, except for the reference stocks and outgroups ($f < 10^{-4}$, $g < 10^{-4}$) and for a main cluster referred to below as cluster I ($f = 1.3 \times 10^{-3}$, $g < 10^{-4}$) and shown in Fig. 1. However, f and g test *P* values were not significant when the strains not included in cluster I were analyzed separately.

Cluster I contained 29 strains, 28 (96.5%) of which carried at least one virulence gene. This cluster also included the two B2 subgroup strains of the ECOR collection and reference strains HB101 and J96. Note that 16 of 19 strains (84%) from our collection carrying the association of *papC*, *hlyC*, and *cnf1* genes were included in this cluster whereas 19 of 20 of our isolates (95%) that carried no virulence genes were found outside the cluster, like the ECOR strains from groups A, B1, and D. Cluster I was identified by the three phylogenetic meth-

TABLE 4. Combination of virulence genes identified in *E. coli* strains isolated from the blood of cancer patients and the feces of healthy volunteers

| Gene combination | % (no.) of positive strains in: | | *P* value[a] |
|---|---|---|---|
| | Blood ($n$ = 155) | Feces ($n$ = 70) | |
| No known virulence factors | 57 (89) | 81.4 (57) | <0.05 |
| *cnf1* + *papC* + *hlyC* | 19 (30) | 4.2 (3) | <0.05 |
| *cnf1* + *papC* | 0.6 (1) | 0 | |
| *cnf1* + *hlyC* | 4 (6) | 0 | |
| *papC* + *hlyC* | 1.2 (2) | 1.4 (2) | NS[b] |
| *cnf1* | 0 | 0 | |
| *papC* | 16 (25) | 11.4 (8) | NS |
| *hlyC* | 1.2 (2) | 0 | |

[a] Chi-square test.
[b] NS, not significant.

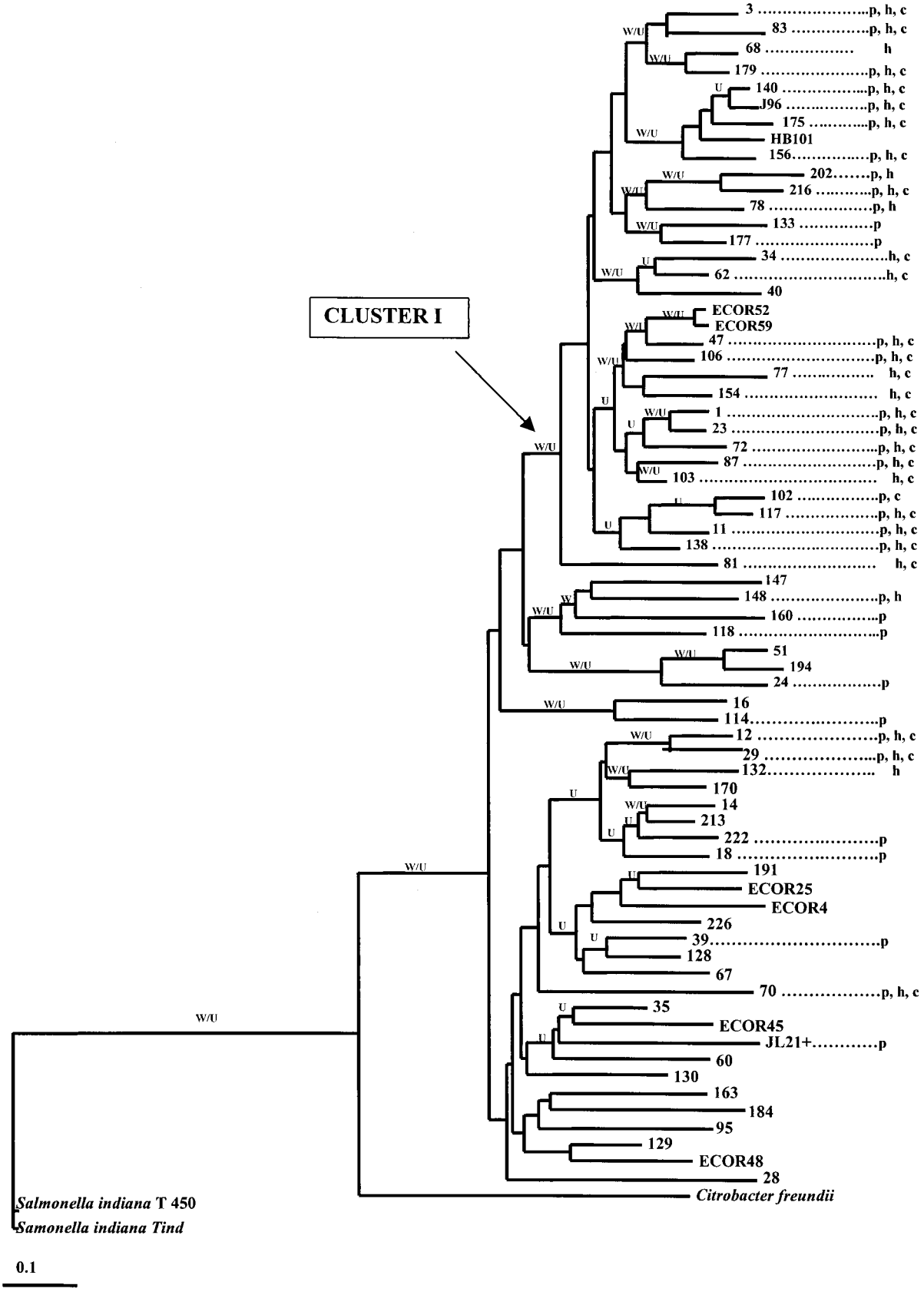


FIG. 1. Rooted neighbor-joining tree generated by the phylogenetic relationships among 61 *E. coli* strains isolated from the blood of cancer patients or the feces of healthy volunteers. The letters W and U indicate branches that were also found by the Wagner and UPGMA methods, respectively. Strains numbered up to 155 were blood isolates from cancer patients, and those with higher numbers were fecal isolates from healthy volunteers. p, h, and c indicates carriage of the *papC*, *hlyC*, and *cnf1* genes, respectively.

ods used. However, the robustness of bootstrap analysis of this cluster was low (0.24).

## DISCUSSION

In this study, we found that the *E. coli* strains isolated from cancer patients with bacteremia carried the virulence-associated *papC*, *hlyC*, and *cnf1* genes more frequently than did strains isolated from the feces of healthy volunteers. This result might be considered somewhat unexpected since cancer patients are often severely immunocompromised and thus could have been infected by much less virulent strains. The frequencies of the *papC*, *hlyC*, *cnf1*, and *iut* genes in both fecal and blood isolates were similar to those reported previously (35, 40). The high frequency of carriage of the *cnf1* gene in our collection of *E. coli* blood isolates was also in agreement with a previous report on the phenotypic expression of CNF in these strains (6), suggesting that *cnf1* plays a role in the pathogenicity of these isolates (16, 20, 27). In the present study, the *iut* gene, which can be either plasmid or chromosome borne (30, 39) and has never been found within a PAI (25), was detected with a non-significantly different frequency in blood and fecal isolates, in contrast with our previous observation concerning aerobactin production in a similar collection of blood isolates from cancer patients (2). However, the prevalence of aerobactin expression we observed in the later collection was dependent on the technique used, since it ranged from 52% using a chemical assay to 89% using a bioassay (2). Other workers reported that the prevalence of the *aero* operon in *E. coli* blood isolates varied from 53% of African strains to 81% of American strains (35). The cause of these variations remains unclear, but their existence makes the role of aerobactin production uncertain in the pathogenesis of *E. coli* bloodstream infections in humans.

Our finding that one or more of the other three virulence genes detected were present in significantly more blood isolates from cancer patients than in fecal isolates from healthy volunteers suggested that they might be important in the pathogenesis of bacteremia. Note that the prevalence of these genes was not significantly different among the strains from bacteremia patients, whatever their underlying disease and immune status, as judged by the number of circulating polymorphonuclear leukocytes, or from the possible urinary source of some cases of bacteremia. This suggests that bacterial factors were of major importance in the pathogenesis of this kind of infection. However, 57% of the bacteremic strains had no virulence factors. It is possible that the corresponding patients had specific immune defects not recorded in this retrospective study. These strains without virulence factors did not group in any specific clone, suggesting that the belonged to the general *E. coli* intestinal populations of the patients.

To establish whether the virulence genes *papC*, *hlyC*, and *cnf1* are linked in the same segment of the chromosome, suggesting their inclusion in a PAI, we hybridized the PFGE patterns generated by three restriction enzymes with specific probes, as did others (9, 54). In 22 of the 32 strains carrying the combination of *papC*, *hlyC*, and *cnf1* genes (69%), these three genes were indeed located on a single restriction fragment, regardless of the restriction enzyme used. The 98% DNA sequence homology between *papC* and *prs* suggested that we had detected *prs* in *E. coli* J96 and that the association of *papC*, *hlyC*, and *cnf1* found in our strains was probably part of a PAI which is very similar, and perhaps identical, to PAI V. No tRNA *pheR* was detected in 20 of 22 strains (90%), carrying the combination of *papC*, *hlyC*, and *cnf1*, suggesting that this combination of genes was probably located in a PAI, such as PAI V. However, in some of our isolates carrying the three virulence genes, the probes hybridized on a band as small as 87 kb, whereas the reported size of PAI V is 110 kb (9, 54), suggesting that one of the three restriction enzymes used had a restriction site(s) in PAI V. This could not be formally confirmed, because the full sequence of PAI V is not available. In some strains carrying two virulence genes, the probes for *papC* and *hlyC* also hybridized on a DNA fragment smaller than 190 kb. This is compatible with the 170-kb long PAI IV, as described in *E. coli* J96 (9). Finally, to our knowledge, the combination of *cnf1* and *hlyC* without *papC*, which we found in six strains, has not been previously reported in a PAI.

To ascertain whether the virulence factors were randomly distributed over the entire spectrum of *E. coli* genotypes, we analyzed the population structure and phylogenetic relationships in a subset of representative *E. coli* isolates. For this purpose, we used the multiprimer RAPD method, which has been successfully applied to the study of the population genetics of several microbial species such as *Candida albicans* (49, 58), *Trypanosoma cruzi* (12, 13a), and *Leishmania* spp. (3). This method generates results equivalent to those of the reference multilocus enzyme electrophoresis method (63) but is less time-consuming and easier to perform. We found that 85% of the *E. coli* strains that carried the combination of *papC*, *hlyC*, and *cnf1* were located in the same genetic cluster, together with the two ECOR subgroup B2 strains. However, the robustness of this cluster is not supported by bootstrap analysis. Thus, although the strains inside the cluster tend to exhibit significantly greater genetic similarity than those outside it, it would hazardous to consider this cluster a real discrete phylogenetic entity.

The cluster included most of the strains that carried at least one virulence gene, whether they were isolated from blood or feces. This observation, combined with the very sporadic occurrence of virulence genes among the strains outside the cluster, suggested that the distribution of virulence genes was not random but was preferentially linked to the cluster.

However, since the cluster also contained the two strains of the ECOR subgroup B2, it probably corresponds to the clonal subgroup B2 of the ECOR collection. In two recent investigations, the *pap*, *hly*, *kps*, and *sfa* genes were detected mainly in the strains from this subgroup (11, 47). Also, a nonrandom distribution of virulence genes was reported among clinical strains isolated from patients with bloodstream infections (35) and neonatal meningitis (5). In a collection of American and African *E. coli* strains isolated from patients with bloodstream infections, *pap*, *hly*, and *sfa* genes were found in a major cluster (35), which was probably identical to the cluster described here. It was recently shown that *E. coli* strains which were isolated from patients with neonatal meningitis and carried the *ibe10* gene encoding brain endothelial cell invasion, the *pap*, *afa*, *sfa*/*foc*, and *hly* genes were distributed mainly in a genetic cluster corresponding to the B2 subgroup of the ECOR collection (5). Thus, whatever their clinical or geographic origin, most strains from this subgroup are likely to carry virulence genes. Since we found that the virulence genes carried by strains from cluster I were likely to be part of PAI IV or V, we suggest that the distribution of PAI IV and V is preferentially linked to strains from the B2 subgroup. Also, since PAIs are suspected to originate from species other than *E. coli* by horizontal transfer and since PAIs are integrated into specific regions of the bacterial chromosome such as the tRNA locus, we suggest that B2 subgroup strains are more prone to integrate PAIs into their bacterial chromosome than are strains outside this subgroup.

As regards the general problem of the *E. coli* population

structure, the strong linkage disequilibrium found in the 60 isolates tested and in cluster I confirms the hypothesis that clonal propagation has a major impact on *E. coli* genotype distribution (44, 52). This linkage disequilibrium persisted even when the genotypes rather than the strains were taken as units of analysis. This indicates that our population of *E. coli* does not correspond to the model of epidemic clonality, as proposed for *Neisseria meningitidis* (36). Our results also do not support the hypothesis that this sample of *E. coli* is divided into discrete genetic entities called discrete typing units (59, 60). The only strong bootstrap value that we found was in fact related to the entire *E. coli* group. All the other branches were weakly supported by bootstrap analysis, which suggests that genetic exchange inhibits strictly separate evolution of the subgroups. This picture is comparable to the picture observed for *T. cruzi*, a parasitic protozoan which is responsible for Chagas' disease and is a typical case of a structured species whose pattern is probably attributable to long-term clonal evolution (57). In this parasite, the discreteness of the genetic lines that subdivide the species is much more apparent than in *E. coli*. *Pseudomonas aeruginosa*, in contrast, exhibits a picture of panmixia, with no apparent linkage disequilibrium (14). The frequency of genetic recombination therefore seems much stronger in this species than in *E. coli*.

In conclusion, the present results suggest that even in immunosuppressed patients with cancer, specific strain characteristics are of major importance in the pathogenesis of *E. coli* bacteremia.

## ACKNOWLEDGMENTS

We thank P. Boquet (Nice, France) for helpful discussions and J. Craig and M. Dreyfus for English revision.

This work was supported in part by Contrat de Recherche Clinique Institut Gustave Roussy 96.1.

## REFERENCES

1. **Andremont, A., R. Lancar, N. A. Lê, J. M. Hattchouel, S. Baron, T. Tavakoli, M. F. Daniel, C. Tancrède, and M. G. Lê.** 1996. Secular trends in mortality associated with bloodstream infections in 4268 patients hospitalized in a cancer referral center between 1975 and 1989. Clin. Microbiol. Infect. **1:**160–167.
2. **Aumont, P., C. Enard, D. Expert, C. Pieddeloup, C. Tancrède, and A. Andremont.** 1989. Production of hemolysin, aerobactin and enterobactin by strains of *Escherichia coli* causing bacteremia in cancer patients, and their resistance to human serum. Res. Microbiol. **140:**21–26.
3. **Banuls, A., F. Guerrini, F. L. Pont, C. Barrera, I. Espinel, R. Guderian, R. Echeverria, and M. Tibayrenc.** 1997. Evidence for hybridization by multilocus enzyme electrophoresis and random amplified polymorphic DNA between *Leishmania braziliensis* and *Leishmania panamensis/guyanensis* in Ecuador. J. Eukaryot. Microbiol. **44:**408–411.
4. **Berg, R.** 1995. Bacterial translocation from the gastrointestinal tract. Trends Microbiol. **3:**149–154.
5. **Bingen, E., B. Picard, N. Brahimi, S. Mathy, P. Desjardins, 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.
6. **Blanco, J., M. Alonso, E. Gonzalez, M. Blanco, and J. Garabal.** 1990. Virulence factors of bacteraemic *Escherichia coli* with particular reference to production of cytotoxic necrotising factor (CNF) by P-fimbriate strains. J. Med. Microbiol. **31:**175–183.
7. **Blanco, J., M. Blanco, M. Alonso, J. Blanco, E. Gonzalez, and J. Garabal.** 1992. Characteristics of haemolytic *Escherichia coli* with particular reference to production of cytotoxic necrotizing factor type 1 (CNF1). Res. Microbiol. **143:**869–878.
8. **Blanco, M., J. Blanco, M. Alonso, and J. Blanco.** 1996. Virulence factors and O groups of *Escherichia coli* isolates from patients with acute pyelonephritis, cystitis and asymptomatic bacteriuria. Eur. J. Epidemiol. **12:**191–198.
9. **Blum, G., V. Falbo, A. Caprioli, and J. Hacker.** 1995. Gene clusters encoding the cytotoxic necrotizing factor type 1, Prs-fimbriae and alpha-hemolysin form the pathogenicity island II of the uropathogenic *Escherichia coli* strain J96. FEMS Microbiol. Lett. **126:**189–195.
10. **Blum, G., M. Ott, A. Lischewski, A. Ritter, H. Imrich, H. Tschape, and J. Hacker.** 1994. Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. Infect. Immun. **62:**606–614.
11. **Boyd, E., and D. Hartl.** 1998. Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution. J. Bacteriol. **180:**1159–1165.
12. **Brisse, S., C. Barnabé, A. Banuls, I. Sidibe, S. Noel, and M. Tibayrenc.** 1998. A phylogenetic analysis of the *Trypanosoma cruzi* genome project CL Brener reference strain by multilocus enzyme electrophoresis and multiprimer random amplified polymorphic DNA fingerprinting. Mol. Biochem. Parasitol. **92:**253–263.
13. **Cherifi, A., M. Contrepois, B. Picard, P. Goullet, J. de Rycke, J. Fairbrother, and J. Barnouin.** 1990. Factors and markers of virulence in *Escherichia coli* from human septicemia. FEMS Microbiol. Lett. **58:**279–283.

13a. **de Lana, M., A. Pinto, C. Barnabé, V. Quesney, S. Noel, and M. Tibayrenc.** 1998. *Trypanosoma cruzi*: compared vectorial transmissibility of three major clonal genotypes by triatoma infestans. Exp. Parasitol. **90:**20–25.

13b. **de Lorenzo, V., A. Bindereif, B. Paw, and J. Neilands.** 1986. Aerobactin biosynthesis and transport genes of plasmid ColV-K30 in *Escherichia coli* K-12. J. Bacteriol. **165:**570–578.

14. **Denamur, E., B. Picard, G. Decoux, J. Denis, and J. Elion.** 1993. The absence of correlation between allozyme and rrn RFLP analysis indicates a high gene flow rate within human clinical *Pseudomonas aeruginosa* isolates. FEMS Microbiol Lett. **110:**275–280.
15. **Falbo, V., T. Pace, L. Picci, E. Pizzi, and A. Caprioli.** 1993. Isolation and nucleotide sequence of the gene encoding cytotoxic necrotizing factor 1 of *Escherichia coli*. Infect. Immun. **61:**4909–4914.
16. **Falzano, L., C. Fiorentini, G. Donelli, E. Michel, C. Kocks, P. Cossart, L. Cabanie, E. Oswald, and P. Boquet.** 1993. Induction of phagocytic behaviour in human epithelial cells by *Escherichia coli* cytotoxic necrotizing factor type 1. Mol. Microbiol. **9:**1247–1254.
17. **Felmlee, T., S. Pellett, and R. A. Welch.** 1985. Nucleotide sequence of an *Escherichia coli* chromosomal hemolysin. J. Bacteriol. **163:**94–105.
18. **Felsenstein, J.** 1985. Confidence limits of phylogenies: an approach using the bootstrap. Evolution **39:**783–791.
19. **Felsenstein, J.** 1993. PHYLIP: Phylogenetic Inference Package, version 3.5c. Department of Genetics, University of Washington, Seattle.
20. **Fiorentini, C., P. Matarrese, E. Straface, L. Falzano, A. Fabbri, G. Donelli, A. Cossarizza, P. Boquet, and W. Malorni.** 1998. Toxin-induced activation of Rho GTP-binding protein increases Bcl-2 expression and influences mitochondrial homeostasis. Exp. Cell Res. **242:**341–350.
21. **Goullet, P., and B. Picard.** 1984. Electrophoretic typing of *Escherichia coli* esterases in septicemia. Presse Med. **13:**1079–1081.
22. **Grothues, D., U. Koopmann, H. van der Hardt, and B. Tümmler.** 1988. Genome fingerprinting of *Pseudomonas aeruginosa* indicates colonization of cystic fibrosis siblings with closely related strains. J. Clin. Microbiol. **26:**1973–1977.
23. **Grothues, D., and B. Tümmler.** 1991. New approaches in genome analysis by pulsed-field gel electrophoresis: application to the analysis of *Pseudomonas* species. Mol. Microbiol. **5:**2763–2776.
24. **Hacker, J., L. Bender, M. Ott, J. Wingender, B. Lund, R. Marre, and W. Goebel.** 1990. Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal *Escherichia coli* isolates. Microb. Pathog. **8:**213–225.
25. **Hacker, J., G. Blum-Oehler, I. Muhldorfer, and H. Tschape.** 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol. Microbiol. **23:**1089–1097.
26. **Herzer, P., S. Inouye, M. Inouye, and T. Whittam.** 1990. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. J. Bacteriol. **172:**6175–6181.
27. **Hofman, P., G. Flatau, E. Selva, M. Gauthier, G. L. Negrate, C. Fiorentini, B. Rossi, and P. Boquet.** 1998. *Escherichia coli* cytotoxic necrotizing factor 1 effaces microvilli and decreases transmigration of polymorphonuclear leukocytes in intestinal T84 epithelial cell monolayers. Infect. Immun. **66:**2494–2500.
28. **Huelsenbeck, J. P., and D. M. Hillis.** 1993. Success of phylogenetic methods in the four-taxon case. Syst. Biol. **42:**247–264.
29. **Hull, R., R. Gill, P. Hsu, B. Minshew, and S. Falkow.** 1981. Construction and expression of recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract infection *Escherichia coli* isolate. Infect. Immun. **33:**933–938.
30. **Jacobson, S., K. Tullus, B. Wretlind, and A. Brauner.** 1988. Aerobactin-mediated uptake of iron by strains of *Escherichia coli* causing acute pyelonephritis and bacteremia. J. Infect. **16:**147–152.
31. **Kim, J.** 1993. Improving of phylogenetic estimation by combining different methods. Syst. Biol. **42:**331–340.
32. **LeBouguenec, C., M. Archambaud, and A. Labigne.** 1992. Rapid and specific detection of the *pap*, *afa*, and *sfa* adhesin-encoding operons in uropathogenic *Escherichia coli* strains by polymerase chain reaction. J. Clin. Microbiol. **30:**1189–1193.
33. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

34. **Mantel, N.** 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. **27:**209–220.
35. **Maslow, J., T. Whittam, C. Gilks, R. Wilson, M. Mulligan, K. Adams, and R. Arbeit.** 1995. Clonal relationships among bloodstream isolates of *Escherichia coli*. Infect. Immun. **63:**2409–2417.
36. **Maynard Smith, J., N. H. Smith, M. O'Rourke, and B. G. Spratt.** 1993. How clonal are bacteria? Proc. Nat. Acad. Sci. USA **90:**4384–4388.
37. **McDaniel, T., K. Jarvis, M. Donnenberg, and J. Kaper.** 1995. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. Proc. Natl. Acad. Sci. USA **92:**1664–1668.
38. **McDaniel, T., and J. Kaper.** 1997. A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. Mol. Microbiol. **23:**399–407.
39. **McDougall, S., and J. Neilands.** 1984. Plasmid-and chromosome-coded aerobactin synthesis in enteric bacteria: insertion sequences flank operon in plasmid-mediated systems. J. Bacteriol. **159:**300–305.
40. **Muhldorfer, I., G. Blum, A. Donohue-Rolfe, H. Heier, T. Olschlager, H. Tschape, U. Wallner, and J. Hacker.** 1996. Characterization of *Escherichia coli* strains isolated from environmental water habitats and from stool samples of healthy volunteers. Res. Microbiol. **147:**625–635.
41. **Ochman, H., and R. Selander.** 1984. Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. **157:**690–693.
42. **Ochman, H., T. Whittam, D. Caugant, and R. Selander.** 1983. Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. J. Gen. Microbiol. **129:**2715–2726.
43. **Opal, S., A. Cross, P. Gemski, and L. 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.
44. **Orskov, F., and I. Orskov.** 1983. From the National Institutes of Health. Summary of a workshop on the clone concept in the epidemiology, taxonomy, and evolution of the enterobacteriaceae and other bacteria. J. Infect. Dis. **48:**346–357.
45. **Oswald, E., J. de Rycke, P. Lintermans, K. van Muylem, J. Mainil, G. Daube, and P. Pohl.** 1991. Virulence factors associated with cytotoxic necrotizing factor type two in bovine diarrheic and septicemic strains of *Escherichia coli*. J. Clin. Microbiol. **29:**2522–2527.
46. **Oswald, E., M. Sugai, A. Labigne, H. C. Wu, C. Fiorentini, P. Boquet, and A. D. O'Brien.** 1994. Cytotoxic necrotizing factor type 2 produced by virulent *Escherichia coli* modifies the small GTP-binding proteins Rho involved in assembly of actin stress fibers. Proc. Natl. Acad. Sci. USA **91:**3814–3818.
47. **Picard, B., J. S. 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.
48. **Picard, B., and P. Goullet.** 1988. Correlation between electrophoretic types B1 and B2 of carboxylesterase B and host-dependent factors in *Escherichia coli* septicaemia. Epidemiol. Infect. **100:**51–61.
49. **Pujol, C., S. Joly, S. Lockhart, S. Noel, M. Tibayrenc, and D. Soll.** 1997. Parity among the randomly amplified polymorphic DNA method, multilocus enzyme electrophoresis, and Southern blot hybridization with the moderately repetitive DNA probe Ca3 for fingerprinting *Candida albicans*. J. Clin. Microbiol. **35:**2348–2358.
50. **Saulnier, P., C. Bourneix, G. Prevost, and A. Andremont.** 1993. Random amplified polymorphic DNA assay is less discriminant than pulsed-field gel electrophoresis for typing strains of methicillin-resistant *Staphylococcus aureus*. J. Clin. Microbiol. **31:**982–985.
51. **Saulnier, P., E. Chachaty, F. Hilali, and A. Andremont.** 1997. Single-step polymerase chain reaction for combined gene detection and epidemiological typing in three bacterial models. FEMS Microbiol. Lett. **150:**311–316.
52. **Selander, R., and B. Levin.** 1980. Genetic diversity and structure in *Escherichia coli* populations. Science **210:**545–547.
53. **Sneath, P. H. A., and R. R. Sokal.** 1973. Numerical taxonomy. The principle and practice of numerical classification. W. H. Freeman & Co., San Francisco, Calif.
54. **Swenson, D., N. Bukanov, D. Berg, and R. Welch.** 1996. Two pathogenicity islands in uropathogenic *Escherichia coli* J96: cosmid cloning and sample sequencing. Infect. Immun. **64:**3736–3743.
55. **Tancrède, C., and A. Andremont.** 1985. Bacterial translocation and gram-negative bacteremia in patients with hematological malignancies. J. Infect. Dis. **152:**99–103.
56. **Tenover, F. C., R. D. Arbeit, R. V. Goering, P. A. Mickelsen, B. E. Murray, D. H. Persing, and B. Swaminathan.** 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J. Clin. Microbiol. **33:**2233–2239.
57. **Tibayrenc, M.** 1995. Population genetics of parasitic protozoa and other microorganisms. Adv. Parasitol. **36:**47–115.
58. **Tibayrenc, M.** 1997. Are *Candida albicans* natural populations subdivided? Trends Microbiol. **5:**253–254.
59. **Tibayrenc, M.** 1998. Beyond strain typing and molecular epidemiology: integrated genetic epidemiology of infectious diseases. Parasitol. Today **14:** 323–329.
60. **Tibayrenc, M.** 1998. Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. Int. J. Parasitol. **28:**85–104.
61. **Tibayrenc, M., F. Kjellberg, J. Arnaud, B. Oury, S. Breniere, M. Darde, and F. Ayala.** 1991. Are eukaryotic microorganisms clonal or sexual? A population genetics vantage. Proc. Natl. Acad. Sci. USA **88:**5129–5133.
62. **Tibayrenc, M., F. Kjellberg, and F. Ayala.** 1990. A clonal theory of parasitic protozoa: the population structures of *Entamoeba*, *Giardia*, *Leishmania*, *Naegleria*, *Plasmodium*, *Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences. Proc. Natl. Acad. Sci. USA **87:**2414–2418.
63. **Tibayrenc, M., K. Neubauer, C. Barnabe, F. Guerrini, D. Skarecky, and F. J. Ayala.** 1993. Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. Proc. Natl. Acad. Sci. USA **90:**1335–1339.
64. **Zingler, G., M. Ott, G. Blum, U. Falkenhagen, G. Naumann, W. Sokolowska-Kohler, and J. Hacker.** 1992. Clonal analysis of *Escherichia coli* serotype O6 strains from urinary tract infections. Microb. Pathog. **12:**299–310.

*Editor:* V. A. Fischetti