

Rapid and Simple Determination of the *Escherichia coli* Phylogenetic Group

OLIVIER CLERMONT, STÉPHANE BONACORSI, AND EDOUARD BINGEN*

Laboratoire d'études de génétique bactérienne dans les infections de l'enfant (EA3105), Université Denis Diderot-Paris 7, Hôpital Robert Debré, Paris, France

Received 15 February 2000/Accepted 7 July 2000

Phylogenetic analysis has shown that *Escherichia coli* is composed of four main phylogenetic groups (A, B1, B2, and D) and that virulent extra-intestinal strains mainly belong to groups B2 and D. Actually, phylogenetic groups can be determined by multilocus enzyme electrophoresis or ribotyping, both of which are complex, time-consuming techniques. We describe a simple and rapid phylogenetic grouping technique based on triplex PCR. The method, which uses a combination of two genes (*chuA* and *yjaA*) and an anonymous DNA fragment, was tested with 230 strains and showed excellent correlation with reference methods.

Escherichia coli is a normal inhabitant of the intestines of most animals, including humans. Some *E. coli* strains can cause a wide variety of intestinal and extra-intestinal diseases, such as diarrhea, urinary tract infections, septicemia, and neonatal meningitis (18). Phylogenetic analyses have shown that *E. coli* strains fall into four main phylogenetic groups (A, B1, B2, and D) (10, 21) and that virulent extra-intestinal strains belong mainly to group B2 and, to a lesser extent, to group D (4, 7, 12, 19), whereas most commensal strains belong to group A. These studies have also given us a better understanding of how pathogenic strains acquire virulence genes (4). Actually, phylogenetic grouping can be done by multilocus enzyme electrophoresis (10, 20) or ribotyping (2–4, 8), but both of these reference techniques are complex and time-consuming and also require a collection of typed strains.

Creation of a subtractive library for two *E. coli* strains belonging to different phylogenetic groups (6) and characterization of an anonymous 14.9-kb fragment strongly associated with neonatal meningitis strains (3) suggested that certain genes or DNA fragments might be specific phylogenetic group markers. Three candidate markers were studied further: (i) *chuA*, a gene required for heme transport in enterohemorrhagic O157:H7 *E. coli* (6, 16, 23, 24); (ii) *yjaA*, a gene initially identified in the recent complete genome sequence of *E. coli* K-12, the function of which is unknown (5); and (iii) an anonymous DNA fragment designated TSPE4.C2 from our subtractive library (6). Here we describe a rapid technique for determining the phylogenetic groups of *E. coli* strains based on PCR detection of the *chuA* and *yjaA* genes and DNA fragment TSPE4.C2. The method was evaluated by testing 230 strains that had already been grouped by using reference methods.

Bacterial strains and growth conditions. The 72 strains of the ECOR collection (17) were kindly provided by R. Selander (Department of Biology, University of Rochester, Rochester, N.Y.). These reference strains, isolated from a variety of hosts and geographic locations, are representative of the range of genotypic variation in the species. Sixty-eight of these strains belong to the four main phylogenetic groups (A, B1, B2, and D), and four are unclassified (10, 21). We also tested a set of 86 *E. coli* strains causing neonatal meningitis (ECNM strains)

(4), 34 *E. coli* strains responsible for neonatal septicemia without meningitis, 30 *E. coli* strains isolated from feces of healthy neonates, the J96 uropathogenic *E. coli* strain (O4:K6) (kindly provided by J. Hacker, Institut für Molekulare Infektionsbiologie, Würzburg, Germany) (22), and 10 verotoxin-producing *E. coli* O157:H7 strains, including 1 strain obtained from A. D. O'Brien (Uniformed Services University of the Health Sciences, Bethesda, Md.) and 9 strains isolated from different locations in France (P. Mariani, Hôpital Robert-Debré, Paris, France). The phylogenetic group distribution of 69 of the 86 ECNM strains has been published previously (4). The other 17 ECNM strains and the 75 remaining clinical isolates were classified by ribotyping as previously described (4). The *E. coli* laboratory K-12 strain MG1655, which belongs to phylogenetic group A, was also studied (10). Bacteria were grown at 37°C in Luria-Bertani broth or on Luria-Bertani agar. When necessary, ampicillin (100 µg per ml) was used.

PCR amplification and Southern blotting. As a first step, PCR was performed with a standard protocol. Each reaction was carried out by using a 20-µl mixture containing 2 µl of 10× buffer (supplied with *Taq* polymerase), 20 pmol of each primer, each deoxynucleoside triphosphate at a concentration of 2 µM, 2.5 U of *Taq* polymerase (ATGC Biotechnologie, Noisy-le-Grand, France), and 200 ng of genomic DNA. The PCR was performed with a Perkin-Elmer GeneAmp 9600 thermal cycler with MicroAm tubes under the following conditions: denaturation for 5 min at 94°C; 30 cycles of 30 s at 94°C, 30 s at 55°C, and 30 s at 72°C; and a final extension step of 7 min at 72°C. The primer pairs used were ChuA.1 (5'-GACGAACCA ACGGTCAGGAT-3') and ChuA.2 (5'-TGCCGCCAGTACC AAAGACA-3'), YjaA.1 (5'-TGAAGTGTTCAGGAGACGCT G-3') and YjaA.2 (5'-ATGGAGAATGCGTTCCTCAAC-3'), and TspE4C2.1 (5'-GAGTAATGTTCGGGGCATTCA-3') and TspE4C2.2 (5'-CGCGCCAACAAAGTATTACG-3'), which generate 279-, 211-, and 152-bp fragments, respectively. In a simplified protocol, a two-step triplex polymerase reaction based on previously described methods (1, 9) was assessed. The components of the reaction mixture were the same as those in the standard protocol, except that (i) DNA was directly provided by 3 µl of bacterial lysate or a piece of a colony, (ii) the six above-mentioned primers were mixed, and (iii) the PCR steps were as follows: denaturation for 4 min at 94°C, 30 cycles of 5 s at 94°C and 10 s at 59°C, and a final extension step of 5 min at 72°C.

Southern blotting was performed by capillary transfer to

* Corresponding author. Mailing address: Service de Microbiologie, Hôpital Robert Debré, 48 Blvd. Sérurier, 75395 Paris cedex 19, France. Phone: 33 1 40 03 23 40. Fax: 33 1 40 03 24 50. E-mail: edouard.bingen@rdb.ap-hop-paris.fr.

TABLE 1. PCR amplification of the *chuA* and *yjaA* genes and DNA fragment TSPE4.C2 in *E. coli* strains from various collections according to phylogenetic group

Strain collection or isolates (no. of strains)	Reference	Group as determined by reference methods (no. of strains) ^a	No. (%) of strains positive for amplification of:		
			<i>chuA</i>	<i>yjaA</i>	TSPE4.C2
ECOR (68)	10	A (25)	0	18 (72) ^b	0
		B1 (16)	0	1 (6) ^c	15 (94) ^c
		D (12)	12 (100)	0	2 (17) ^d
		B2 (15)	15 (100)	15 (100)	12 (80) ^e
Neonatal meningitis (86)	4	A (5)	0	5 (100)	0
		B1 (3)	0	1 (33)	2 (66)
		D (18)	18 (100)	0	2 (11)
		B2 (60)	60 (100)	60 (100)	59 (98)
Verotoxin-producing <i>E. coli</i> O157:H7 (10)	This study	D (10)	10 (100)	0	0
Other clinical strains (64)	This study	A (12)	0	9 (75)	0
		B1 (4)	0	0	4 (100)
		D (11)	11 (100)	0	1 (9)
		B2 (37)	37 (100)	37 (100)	34 (92)
<i>E. coli</i> K-12	10	A (1)	0	1	0
<i>E. coli</i> J96	This study	B2 (1)	1	1	0
All strains (230)		A (43)	0	33 (77)	0
		B1 (23)	0	2 (9)	21 (91)
		D (51)	51 (100)	0	5 (10)
		B2 (113)	113 (100)	113 (100)	105 (93)

^a Phylogenetic groups were determined by multilocus enzyme electrophoresis or ribotyping.

^b The strains negative for amplification of *yjaA* were ECOR 4, ECOR 15, ECOR 16, ECOR 17, ECOR 22, ECOR 23, and ECOR 24.

^c The strain negative for amplification of *yjaA* and positive for amplification of TSPE4.C2 was ECOR 70.

^d The strains positive for amplification of TSPE4.C2 were ECOR 49 and ECOR 50.

^e The strains negative for amplification of TSPE4.C2 were ECOR 53, ECOR 59, and ECOR 60.

positively charged nylon membranes. Hybridization was performed at 65°C in 1% sodium dodecyl sulfate–1 M NaCl–50 mM Tris HCl (pH 7.5)–1% blocking reagent (Boehringer, Mannheim, Germany). The membranes were washed in 2× SSC for 15 min at room temperature, then in 2× SSC–0.1% sodium dodecyl sulfate for 30 min at 65°C, and finally in 0.1× SSC for 5 min at room temperature (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate). Chemiluminescence detection was performed according to the manufacturer's instructions (DIG Luminescent Detection Kit for nucleic acids; Boehringer). The probes were produced by PCR according to the manufacturer's instructions (PCR DIG Probe Synthesis Kit; Boehringer) by using the primers and amplification procedure described above for the standard protocol.

PCR grouping results. A total of 230 strains were analyzed. Their phylogenetic groups, as determined by reference methods, were as follows: 43 belonged to group A, 23 belonged to group B1, 51 belonged to group D, and 113 belonged to group B2. Table 1 shows the PCR results for the entire set of strains according to their phylogenetic groups. The *chuA* gene was present in all strains belonging to groups B2 and D and was absent from all strains belonging to groups A and B1. This allowed us to separate groups B2 and D from groups A and B1. In the same way, the *yjaA* gene allowed perfect discrimination between group B2 (100% of the strains were positive) and group D (100% of the strains were negative). Finally, clone TSPE4.C2 was present in all but two of the group B1 strains and absent from all group A strains. All the PCR results were confirmed by Southern hybridization (data not shown). The results of these three amplifications made it possible to establish a dichotomous decision tree (Fig. 1) for phylogenetic

grouping. With this tree, 228 of the 230 strains tested (99%) were correctly grouped, while only 2 strains were wrongly classified (group B1 strains were identified as group A strains). Identical results were obtained with standard and simplified PCR protocols. Figure 2 shows the different profiles obtained by triplex PCR for the four phylogenetic groups.

In this work, we developed a PCR method to rapidly deter-

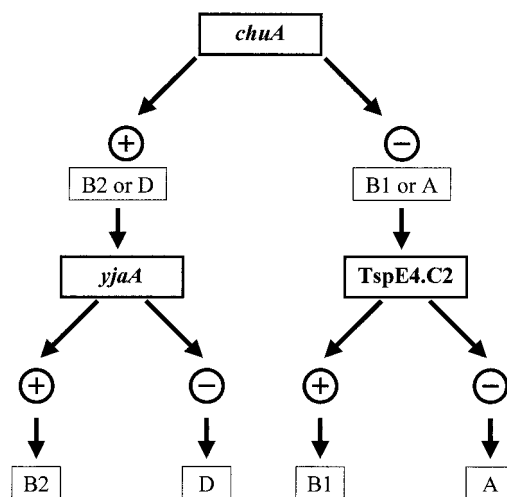


FIG. 1. Dichotomous decision tree to determine the phylogenetic group of an *E. coli* strain by using the results of PCR amplification of the *chuA* and *yjaA* genes and DNA fragment TSPE4.C2.

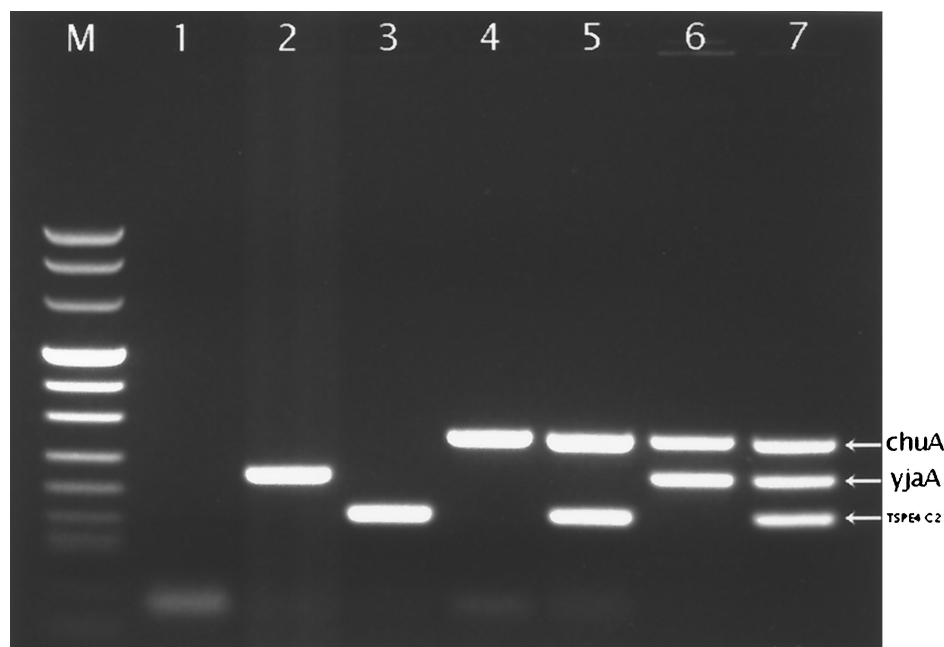


FIG. 2. Triplex PCR profiles specific for *E. coli* phylogenetic groups. Each combination of *chuA* and *yjaA* gene and DNA fragment TSPE4.C2 amplification allowed phylogenetic group determination of a strain. Lanes 1 and 2, group A; lane 3, group B1; lanes 4 and 5, group D; lanes 6 and 7, group B2. Lane M contained markers.

mine the phylogenetic groups of *E. coli* strains and obtained an accuracy of more than 99% compared to the reference method. Phylogenetic characterization of *E. coli* strains on the basis of a very few phenotypic or genotypic features initially appeared to be very difficult. Such genotypic traits (presence or absence of a gene, for example) must meet different criteria for use in phylogenetic characterization. First, the gene must have been acquired or deleted when the group that it characterizes emerged. Second, the same gene must have been “stabilized,” thereby ruling out its subsequent deletion or horizontal transfer among bacteria belonging to other phylogenetic groups. Finally, recombination events in the candidate gene must be very rare. In other words, the gene product must not be targeted by natural selection, which favors new genetic recombinations (24). Previous attempts to identify specific phylogenetic group characteristics based on phenotypic (21) or genotypic features (11) were not sufficiently discriminative. For the first time, we describe the use of two genes and an anonymous DNA fragment in a simple phylogenetic grouping method. Too little information is available on *yjaA* and the DNA fragment to speculate on their evolutionary history. In contrast, the study by Wyckoff et al. (25) of heme transport genes, together with our results, suggests that *chuA* was acquired by sister groups B2 and D (15) soon after their emergence rather than being present in a common ancestor and subsequently being lost by groups B1 and A.

However, two strains (ECOR 70 and an ECNM strain) belonging to phylogenetic group B1 were classified in group A by our method. This discrepancy may be explained by an intermediate genetic base between these two groups in these strains and by the fact that the region studied with our method (*chuA*, *yjaA*, and TSPE4.C2 are located at 78.7 min [25], 90.8 min [5], and approximately 87 min [6], respectively, in relation to the genome of *E. coli* K-12) may be more closely related to group A regions than to the regions studied with the reference methods. Indeed, it has been demonstrated that groups A and B1 are sister groups (15). Moreover, recent multiple chromosomal

nucleotide sequence analysis has shown that ECOR 70 may be considered a “hybrid” strain, in which some housekeeping genes exhibit nucleotide sequences shared by group A ECOR strains and some other genes exhibit nucleotide sequences shared by group B1 ECOR strains (15; E. Denamur, personal communication). Thus, the phylogenetic group of ECOR 70 remains to be settled. In addition to the rapidity of our PCR-based method, no reference collection is required, meaning that the assay can easily be used in any laboratory. Furthermore, in contrast to other methods, group allocation is unequivocal. Indeed, the four unclassified strains in the ECOR collection, ECOR 31, ECOR 37, ECOR 42, and ECOR 43, were classified by our method; the first three strains belong to group D, and the fourth belongs to group A. It is noteworthy that all the sequences of the housekeeping genes studied in the latter strain were characteristic of those found in group A strains (Denamur, personal communication).

In conclusion, our simple and rapid phylogenetic grouping technique could have several practical applications. The first is in bioclinical practice, given the established link between phylogenetic group and virulence (4, 7, 12, 19). The second is as a biotechnological screening tool for eliminating potentially pathogenic strains when candidate strains for cloning are screened. Such screening tools have been developed, for example, to identify *E. coli* K-12 strains by PCR (13) or to detect *E. coli* strains with no known virulence genes by a reverse dot blot procedure described by Kuhnert et al. (14). Our method has the advantage of being capable of identifying nonpathogenic strains other than *E. coli* K-12 and suitable for large-scale strain screening. Thus, after all strains belonging to groups B2 and D, which are potentially pathogenic, are eliminated, the reverse dot blot technique (14) could be applied to group A or B1 strains to eliminate rare strains harboring virulence factors.

We thank Walter Nanni and Colin Tinsley for technical help and Xavier Nassif, Jacques Elion, and Erick Denamur for helpful discussions.

This work was supported in part by the Programme de Recherche Fondamentale en Microbiologie, Maladies Infectieuses et Parasitaires (Appel d'offre 1998), "Recherche de déterminants génétiques de pathogénicité chez *E. coli* K1 responsable de méningite néonatale," and by the Programme Hospitalier de Recherche Clinique (grant AOM 96069).

REFERENCES

1. Belgrader, P., W. Benett, D. Hadley, G. Long, R. Mariella, Jr., F. Milanovich, S. Nasarabadi, W. Nelson, J. Richards, and P. Stratton. 1998. Rapid pathogen detection using a microchip PCR array instrument. *Clin. Chem.* **44**: 2191–2194.
2. Bingen, E., E. Denamur, and J. Elion. 1994. Use of ribotyping in epidemiological surveillance of nosocomial outbreaks. *Clin. Microbiol. Rev.* **7**:311–317.
3. Bingen, E., E. Denamur, N. Brahimi, and J. Elion. 1996. Genotyping may provide rapid identification of *Escherichia coli* K1 organisms that cause neonatal meningitis. *Clin. Infect. Dis.* **22**:152–156.
4. 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.
5. Blattner, F. R., G. I. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**:1453–1461.
6. Bonacorsi, S. P. P., O. Clermont, C. Tinsley, I. Le Gall, J. C. Beaudoin, J. Elion, X. Nassif, and E. Bingen. 2000. Identification of regions of the *Escherichia coli* chromosome specific for neonatal meningitis-associated strains. *Infect. Immun.* **68**:2096–2101.
7. Boyd, E. F., and D. L. Hartl. 1998. Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution. *J. Bacteriol.* **180**:1159–1165.
8. Desjardins, 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 the restriction fragment length polymorphism. *J. Mol. Evol.* **41**:440–448.
9. Haedicke, W., H. Wolf, W. Ehret, and U. Reischl. 1996. Specific and sensitive two-step polymerase reaction assay for the detection of the *Salmonella* species. *Eur. J. Clin. Microbiol. Infect. Dis.* **15**:603–607.
10. 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.
11. Hurtado, A., and F. Rodriguez-Valera. 1999. Accessory DNA in the genomes of representatives of the *Escherichia coli* reference collection. *J. Bacteriol.* **181**:2548–2554.
12. Johnson, J. R., and A. L. Stell. 2000. Extended virulence genotypes of *Escherichia coli* strains from patients with uresepsis in relation to phylogeny and host compromise. *J. Infect. Dis.* **181**:261–272.
13. Kuhnert, P., J. Nicolet, and J. Frey. 1995. Rapid and accurate identification of *Escherichia coli* K-12 strains. *Appl. Environ. Microbiol.* **61**:4135–4139.
14. Kuhnert, P., J. Hacker, I. Mühlendorfer, A. P. Burnens, J. Nicolet, and J. Frey. 1997. Detection system for *Escherichia coli*-specific virulence genes: absence of virulence determinants in B and C strains. *Appl. Environ. Microbiol.* **63**:703–709.
15. Lecointre, G., L. Rachdi, P. Darlu, and E. Denamur. 1998. *Escherichia coli* molecular phylogeny using the incongruence length difference test. *Mol. Biol. Evol.* **15**:1685–1695.
16. Mills, M., and S. Payne. 1995. Genetics and regulation of haem iron transport in *Shigella dysenteriae* and detection of an analogous system in *Escherichia coli* O157:H7. *J. Bacteriol.* **177**:3004–3009.
17. Ochman, H., and R. K. Selander. 1984. Standard reference strains of *Escherichia coli* from natural populations. *J. Bacteriol.* **157**:690–692.
18. Orskov, F., and I. Orskov. 1992. *Escherichia coli* serotyping and disease in man and animals. *Can. J. Microbiol.* **38**:699–704.
19. 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.
20. Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour, and T. S. Whittam. 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl. Environ. Microbiol.* **51**:873–884.
21. Selander, R. K., D. A. Caugant, and T. S. Whittam. 1987. Genetic structure and variation in natural populations of *Escherichia coli*, p. 1625–1648. In F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
22. Swenson, D. L., N. O. Bukanov, D. E. Berg, and R. A. Welch. 1996. Two pathogenicity islands in uropathogenic *Escherichia coli* J96: cosmid cloning and sample sequencing. *Infect. Immun.* **64**:3736–3743.
23. Torres, A., and S. Payne. 1997. Haem iron-transport system in enterohaemorrhagic *Escherichia coli* O157:H7. *Mol. Microbiol.* **23**:825–833.
24. Whittam, T. 1996. Genetic variation and evolutionary processes in natural populations of *Escherichia coli*, p. 2708–2720. In F. C. Neidhardt (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
25. Wyckoff, E. E., D. Duncan, A. G. Torres, M. Mills, K. Maase, and S. M. Payne. 1998. Structure of the *Shigella dysenteriae* haem transport locus and its phylogenetic distribution in enteric bacteria. *Mol. Microbiol.* **28**:1139–1152.