J. Microbiol. Biotechnol. (2005), **15**(3), 672–677

JOURNAL OF MICROBIOLOGY AND BIOTECHNOLOGY

Diversity and Genotypic Structure of ECOR Collection Determined by Repetitive Extragenic Palindromic PCR Genome Fingerprinting

HWANG, KEUM-OK, HYO-MI JANG, AND JAE-CHANG CHO*

Department of Environmental Sciences, Hankuk University of Foreign Studies, Yong-In Shi, Kyung-Ki Do 449-791, Korea

Received: October 31, 2004
Accepted: January 5, 2005

Abstract The standard reference collection of strains for *E. coli*, the ECOR collection, was analyzed by a genome-based typing method. Seventy-one ECOR strains were subjected to repetitive extragenic palindromic PCR genome fingerprinting with BOX primers (BOX-PCR). Using a similarity value of 0.8 or more after cluster analysis of BOX-PCR fingerprinting patterns to define the same genotypes, we identified 28 genotypes in the ECOR collection. Shannon's entropy-based diversity index was 3.07, and the incident-based coverage estimator indicated potentially 420 genotypes among *E. coli* populations. Chi-square test of goodness-of-fit showed statistically significant association between the genotypes defined by BOX-PCR fingerprinting and the groups previously defined by multi-locus enzyme electrophoresis. This study suggests that the diversification of *E. coli* strains in natural populations is actively ongoing, and rep-PCR fingerprinting is a convenient and reliable method to type *E. coli* strains for the purposes ranging from ecology to quarantine.

Key words: Genotype, diversity, ECOR, BOX-PCR

Escherichia coli, a species of Gram-negative bacteria, is a common member of normal microbial flora of mammals, and the species consists of harmless commensal strains as well as pathogenic strains. Since the pathogenic strains can cause a variety of intestinal and extra-intestinal diseases, such as diarrhea, urinary tract infections, septicemia, and neonatal meningitis, resulting in severe sickness and mortality, the groupings of *E. coli* strains have mainly been made by human medical syndromes, which include enterohemorrhagic (EHEC), enteropathogenic (EPEC), and enterotoxigenic (ETEC) groups. Furthermore, the molecular typing methods for *E. coli* strains were developed based on O:H serotypes, virulence, and adherence properties, especially for clinical purposes [1, 9, 26, 36]. However, little emphasis has been given to the study of genetic diversity and structure of *E. coli* in natural populations.

The first attempt to assess the genetic structure and diversity of *E. coli* was carried out with 829 isolates obtained mainly from humans [24]. Later in 1984, the standard reference collection of strains for *E. coli*, the ECOR collection, was established [27]. The ECOR collection comprised of 72 strains isolated from human or nonhuman mammalian orders of a variety of geographic origins. After the establishment of this standard reference collection of *E. coli* strains, many studies [32, 37, 38] extended the first Milkman's work [24], and bacterial population genetics significantly benefited from the studies [3, 4, 12, 15, 16, 17, 25, 30]. However, all the previous studies on the ECOR strains were performed using multi-locus enzyme electrophoresis (MLEE) [31], which has long been a standard method in eukaryotic population genetics. Although the MLEE analysis, which is based on electrophoretic mobility of enzymes encoded in several loci, has served as a standard method in the field of population biology, the resolution of the method as a molecular typing tool is considered to be far lower than the resolutions achieved by DNA-sequences-based methods in the genomics era. In this study, we applied repetitive extragenic palindromic (rep) PCR genome fingerprinting technique [7] to fully resolve the ECOR strains with the resolution of genome level.

The 72 *E. coli* strains of the ECOR collection [27], which includes isolates from a wide variety of hosts and geographic regions (Table 1), were provided by T. S. Whittam (Michigan State University, East Lansing, MI, U.S.A.). The *E. coli* strains were grown on Luria-Bertani (LB) media at 37°C and stored at −80°C in 15% glycerol. Repetitive extragenic palindromic PCR (rep-PCR) genome fingerprinting of the ECOR strains was carried out with a BOX-A1R primer (hereafter, BOX-PCR) according to the protocol of Cho and Tiedje [7]. DNA solutions prepared by the boiling methods [35] were used for PCR templates. Seventy-one

*Corresponding author
Phone: 82-31-330-4350; Fax: 82-31-330-4529;
E-mail: chojc@hufs.ac.kr

Table 1. List of ECOR strains.

Strain no.	MLEE group	Host	Geographic origin	Strain no.	MLEE group	Host	Geographic origin
ECOR-01	A	Human	USA	ECOR-37	E	Marmoset	USA
ECOR-02	A	Human	USA	ECOR-38	D	Human	USA
ECOR-03	A	Dog	USA	ECOR-39	D	Human	Sweden
ECOR-04	A	Human	USA	ECOR-40	D	Human	Sweden
ECOR-05	A	Human	USA	ECOR-41	D	Human	Tonga
ECOR-06	A	Human	USA	ECOR-42	E	Human	USA
ECOR-07	A	Orangutan	USA	ECOR-43	E	Human	Sweden
ECOR-08	A	Human	USA	ECOR-44	D	Cougar	USA
ECOR-09	A	Human	Sweden	ECOR-45	B1	Pig	Indonesia
ECOR-10	A	Human	Sweden	ECOR-46	D	Ape	USA
ECOR-11	A	Human	Sweden	ECOR-47	D	Sheep	New Guinea
ECOR-12	A	Human	Sweden	ECOR-48	D	Human	Sweden
ECOR-13	A	Human	Sweden	ECOR-49	D	Human	Sweden
ECOR-14	A	Human	Sweden	ECOR-50	D	Human	Sweden
ECOR-15	A	Human	Sweden	ECOR-51	B2	Human	USA
ECOR-16	A	Leopard	USA	ECOR-52	B2	Orangutan	USA
ECOR-17	A	Pig	Indonesia	ECOR-53	B2	Human	USA
ECOR-18	A	Celebese ape	USA	ECOR-54	B2	Human	USA
ECOR-19	A	Celebese ape	USA	ECOR-55	B2	Human	Sweden
ECOR-20	A	Steer	Bali	ECOR-56	B2	Human	Sweden
ECOR-21	A	Steer	Bali	ECOR-57	B2	Gorilla	USA
ECOR-22	A	Steer	Bali	ECOR-58	B1	Lion	USA
ECOR-23	A	Elephant	USA	ECOR-59	B2	Human	USA
ECOR-24	A	Human	Sweden	ECOR-60	B2	Human	Sweden
ECOR-25	A	Dog	USA	ECOR-61	B2	Human	Sweden
ECOR-26	B1	Human	USA	ECOR-62	B2	Human	Sweden
ECOR-27	B1	Giraffe	USA	ECOR-63	B2	Human	Sweden
ECOR-28	B1	Human	USA	ECOR-64	B2	Human	Sweden
ECOR-29	B1	Kangaroo rat	USA	ECOR-65	B2	Celebese ape	USA
ECOR-30	B1	Bison	Canada	ECOR-66	B1	Celebese ape	USA
ECOR-31	E	Leopard	USA	ECOR-67	B1	Goat	Indonesia
ECOR-32	B1	Giraffe	USA	ECOR-68	B1	Giraffe	USA
ECOR-33	B1	Sheep	USA	ECOR-69	B1	Celebese ape	USA
ECOR-34	B1	Dog	USA	ECOR-70	B1	Gorilla	USA
ECOR-35	D	Human	USA	ECOR-71	B1	Human	Sweden
ECOR-36	D	Human	USA	ECOR-72	B1	Human	Sweden

ECOR strains were subjected to the rep-PCR genome fingerprinting, since one strain was missing due to contamination during subculture. PCR products were resolved on 15-cm-long 1.5% agarose gels, and a 1-kb DNA size ladder was used at both ends in the middle of the gels. The gels were stained with ethidium bromide, and gel images were digitized using a gel documentation system equipped with a charge-coupled device video camera and stored as TIFF files. These digitized images were converted, normalized with the above-mentioned DNA size markers, and analyzed with GelCompar II software (Applied Maths, Kortrijk, Belgium). The rolling-disk background subtraction method was applied. For BOX-PCR fingerprint analysis, similarity matrices of whole densitometric curves of the gel tracks were calculated by using the pairwise Pearson's product-moment correlation coefficient (r value) [21] as recommended by Rademaker *et al.* [29]. It is insensitive to the relative concentration of bands between fingerprints, discontinuous noise, and overall intensity of fingerprint.

Cluster analyses of similarity matrices were performed by the unweighted pair group method using arithmetic average (UPGMA) [21]. To analyze the reproducibility of BOX-PCR fingerprinting, several strains were subjected several times to BOX-PCR. A comparison of the resulting fingerprinting patterns resolved on the independent gels yielded similarity coefficient (r value) of >0.8, which is consistent with those from other studies using rep-PCR [7, 28, 29]. Hence, we chose a similarity value of 0.8 or more to indicate strains of the same genotype, resulting in 28 genotypes (Fig. 1).

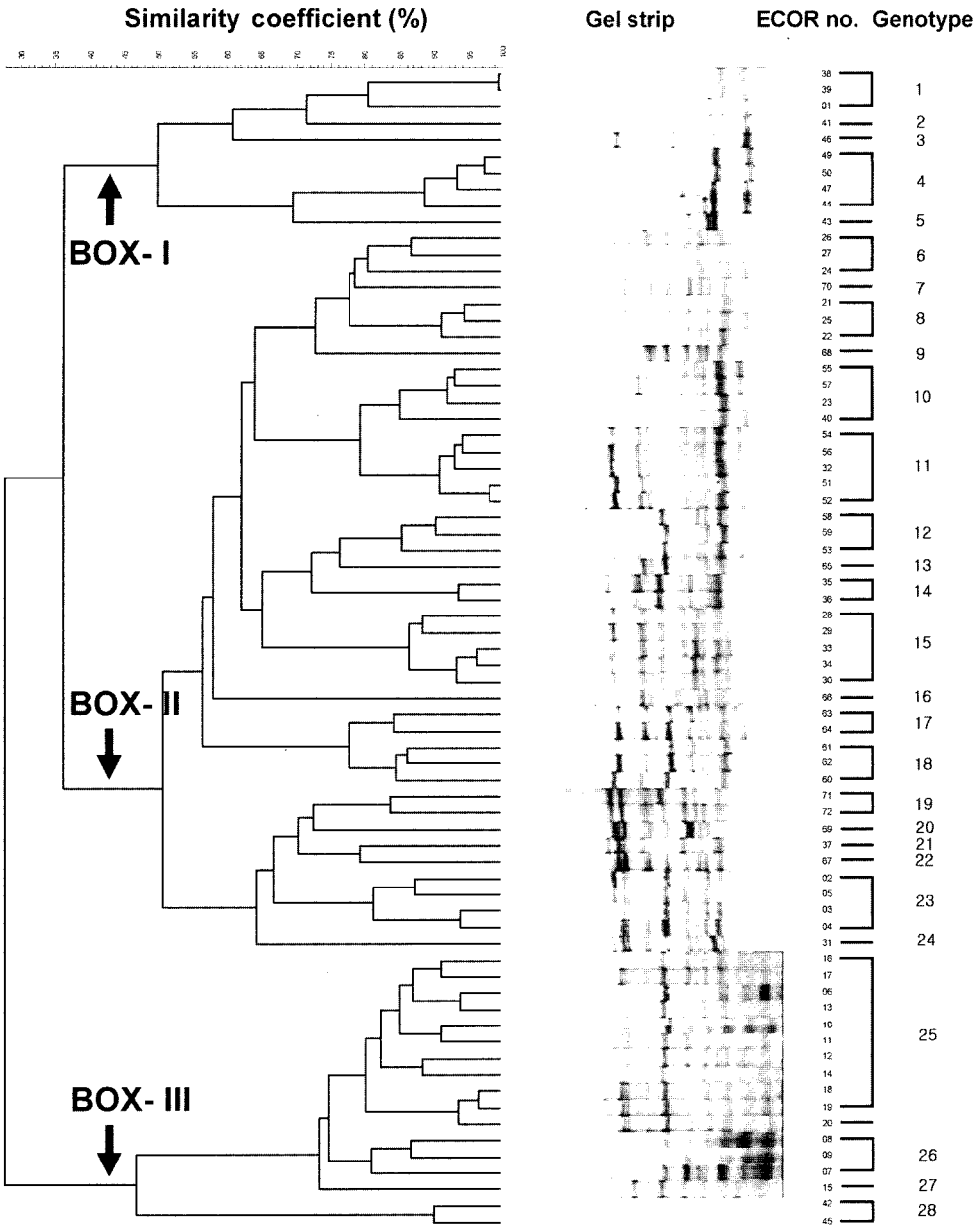

Fig. 1. rep-PCR whole genome fingerprinting of ECOR strains. A similarity dendrogram was constructed using the unweighted pair group method with arithmetic mean (UPGMA).
Scale bar indicates similarity coefficient (%), converted from Pearson correlation coefficient.

Three major clusters, BOX-I, BOX-II, and BOXIII, were found and they comprised members mainly from MLEE group D, B1/B2, and A, respectively. Chi-square test of goodness-of-fit showed statistically significant association between the genotypes defined by BOX-PCR fingerprinting and the groups previously defined by MLEE (C^2=236.57, p<0.005, df=108). However, a little incongruence between the two methods was observed. Six genotypes, 1, 6, 10, 11, 12, and 28, consisted of strains belonging to different MLEE groups. Similarly, Clermont *et al.* [8] observed the incongruence between the methods. When analyzed by triplex PCR, ECOR strains 23 and 24 in genotypes 10 and 6, which were defined as MLEE group A, showed PCR amplification results different from the other MLEE group A strains. Besides this, ECOR strains 53 and 59 in genotype 12, which were defined as MLEE group B2, showed PCR amplification results different from other MLEE group B2 strains. The observed incongruence between the methods was attributed to the differences in testing materials of the methods.

While the MLEE method analyzes the electrophoretic mobility of several proteins, the BOX-PCR method analyzes the whole genome sequence. Compared to the phylogeny of single or several gene(s) or protein(s), a whole-genome-based method can give more reliable measures for the genetic distance and relatedness, since it compares the overall genome sequences, hence is at least not biased (e.g., different substitution rate between genes, horizontal gene transfer). The best measurement for explaining genetic relationship is whole genomic DNA-DNA hybridization, which is the only official method for describing bacterial species [6, 34]. However, whole genomic DNA-DNA hybridization is laborious and time-consuming, thus not appropriate for studies dealing with many strains.

The rep-PCR genome fingerprinting method is based on the gel-profiling of PCR-amplified DNA sequences between short repetitive sequences dispersed on the bacterial genome. Three families of the short repetitive sequences, REP [11, 12], ERIC [14, 33], and BOX [18, 23], have the potential to form stem-loop structures and play an important role in the bacterial genome organization [19, 22], which is considered to be under selective pressure. Consequently, the dispersion pattern of the REP, ERIC, and BOX sequences indicates the structure and evolution of the bacterial genome [22]. Rademaker *et al.* [28] have shown that the similarity coefficient calculated from electrophoretic patterns of rep-PCR amplicons correlated well with whole genomic DNA-DNA similarity, and that the similarity coefficient of 80%, which was the cut-off level to define the genotypes, corresponded to the whole genomic DNA-DNA similarity of 95%. Hence, the cut-off value used in this study provided enough resolution to differentiate ECOR strains at the genome organization level.

Cluster analysis resulted in a total of 28 *E. coli* genotypes out of 71 strains in the ECOR collection. When incident-based coverage estimator (ICE) [5] was applied, we estimated potentially 420 genotypes in the species *E. coli*. Diversity index calculated using Shannon's entropy (*H*) for the whole ECOR collection was 3.07. MLEE group B1 showed the highest diversity (H=2.19) and MLEE group D showed the lowest diversity (H=1.65).

Considering that active diversification results in higher diversity, the diversity indices estimated are consistent with an evolutionary scenario of species *E. coli*. Lecointre *et al.* [20] suggested an evolutionary sequence that MLEE groups D and B2 are ancient *E. coli* groups, followed by the sister MLEE groups A and B1. Anton *et al.* [2] also reported similar findings, but concluded that highly homogeneous MLEE group B2 might be the most ancient group. The highest genotypic diversity of MLEE group B1 indicates that the diversification of this group is actively ongoing, implying high frequency of recombination with DNAs from related strains. Interestingly, most of the EPEC and EHEC strains were reported to be members of MLEE group B1, when analyzed by the mononucleotide repeated (MNR) loci method [10]. On the other hand, MLEE group A, which includes mainly nonpathogenic commensal strains, showed relatively low diversity (H=1.68). The low diversity of this rapidly evolving group might be explained by the hypothesis that the ecological niches of MLEE group A populations are more homogeneous than other MLEE group members, resulting in low level of niche differentiation and diversification. However, such interpretation is based on the circumstantial evidences and out of the scope of this study.

In conclusion, we analyzed standard *E. coli* strains (ECOR collection) using rep-PCR genome fingerprinting to differentiate the strains at the genotypic level, and 28 genotypes were identified. High genotypic diversity was observed for whole ECOR strains, and MLEE group B1 showed the highest diversity among the MLEE groups, indicating that diversification is actively ongoing. rep-PCR fingerprinting was suggested to be a convenient and reliable method to type *E. coli* strains for ecological, clinical, as well as quarantine purposes. To the best of our knowledge, the rep-PCR genotyping of ECOR strains in this study is the first genome-based study on the ECOR collection, providing insight into the ecological and genetic structure of *E. coli* in natural populations.

Acknowledgments

We thank T. S. Whittam (Michigan State University) for providing ECOR strains and K.-S. Hwang (Ewha Womans University) for GelCompar II software access. This study was supported by the Korea Research Foundation Grant (KRF-2002-015-CS0057).

REFERENCES

1. Amhaz, J. M., A. Andrade, S. Y. Bando, T. L. Tanaka, C. A. Moreira-Filho, and M. B. Martinez. 2004. Molecular typing and phylogenetic analysis of enteroinvasive *Escherichia coli* using the fliC gene sequence. *FEMS Microbiol. Lett.* **235:** 259–264.
2. Anton, A. I., A. J. Martinez-Murcia, and F. Rodriguez-Valera. 1999. Intraspecific diversity of the 23S rRNA gene and the spacer region downstream in *Escherichia coli*. *J. Bacteriol.* **181:** 2703–2709.
3. Bae, J. W., J. J. Kim, C. O. Jeon, K. Kim, J. J. Song, S. G. Lee, H. Poo, C. M. Jung, Y. H. Park, and M. H. Sung. 2003. Application of denaturing gradient gel electrophoresis to estimate the diversity of commensal thermophiles. *J. Microbiol. Biotechnol.* **13:** 1008–1012.
4. Beyene, D., S. Kassa, F. Ampy, A. Asseffa, T. Gebremedhin, and P. van Berkum. 2004. Ethiopian soils harbor natural populations of rhizobia that form symbioses with common

bean (*Phaseolus vulgaris* L.). *Arch. Microbiol.* **181:** 129–136.
5. Chao, A. and S.-M. Lee. 1992. Estimating the number of classes via sample coverage. *J. Am. Stat. Assoc.* **87:** 210–217.
6. Cho, J. C. and J. M. Tiedje. 2001. Bacterial species determination from DNA-DNA hybridization by using genome fragments and DNA microarrays. *Appl. Environ. Microbiol.* **67:** 3677–3682.
7. Cho, J. C. and J. M. Tiedje. 2000. Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. *Appl. Environ. Microbiol.* **66:** 5448–5456.
8. Clermont, O., S. Bonacorsi, and E. Bingen. 2000. Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl. Environ. Microbiol.* **66:** 4555–4558.
9. De Baets, L., I. Van der Taelen, M. De Filette, D. Pierard, L. Allison, H. De Greve, J. P. Hernalsteens, and H. Imberechts. 2004. Genetic typing of shiga toxin 2 variants of *Escherichia coli* by PCR-restriction fragment length polymorphism analysis. *Appl. Environ. Microbiol.* **70:** 6309–6314.
10. Diamant, E., Y. Palti, R. Gur-Arie, H. Cohen, E. M. Hallerman, and Y. Kashi. 2004. Phylogeny and strain typing of *Escherichia coli*, inferred from variation at mononucleotide repeat loci. *Appl. Environ. Microbiol.* **70:** 2464–2473.
11. Gilson, E., J. M. Clement, D. Brutlag, and M. Hofnung. 1984. A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. *EMBO J.* **3:** 1417–1421.
12. Higgins, C. F., G. F. Ames, W. M. Barnes, J. M. Clement, and M. Hofnung. 1982. A novel intercistronic regulatory element of prokaryotic operons. *Nature* **298:** 760–762.
13. Hohwy, J., J. Reinholdt, and M. Kilian. 2001. Population dynamics of *Streptococcus mitis* in its natural habitat. *Infect. Immun.* **69:** 6055–6063.
14. Hulton, C. S., C. F. Higgins, and P. M. Sharp. 1991. ERIC sequences: A novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. *Mol. Microbiol.* **5:** 825–834.
15. Kim, S. H., K. B. Lee, J. S. Lee, and Y. H. Cho. 2003. Genomic diversification by phage-derived genomic islands in *Pseudomonas aeruginosa*. *J. Microbiol. Biotechnol.* **13:** 783–832.
16. Kim, S. O., O. Y. Lim, S. H. Kim, J. Y. Kim, Y. H. Kang, and B. K. Lee. 2003. Pulsed-field gel electrophoresis and mutation typing of gyrA gene of quinolone-resistant *Salmonella enterica* serovar paratyphi A isolated from outbreak and sporadic cases, 1998–2002, Korea. *J. Microbiol. Biotechnol.* **13:** 155–158.
17. Kim, T. W., S. H. Jung, J. Y. Lee, S. K. Choi, S. H. Park, J. S. Jo, and H. Y. Kim. 2003. Identification of lactic acid bacteria in kimchi using SDS-PAGE profiles of whole cell proteins. *J. Microbiol. Biotechnol.* **13:** 119–124.
18. Koeuth, T., J. Versalovic, and J. R. Lupski. 1995. Differential subsequence conservation of interspersed repetitive *Streptococcus pneumoniae* BOX elements in diverse bacteria. *Genome Res.* **5:** 408–418.
19. Krawiec, S. and M. Riley. 1990. Organization of the bacterial chromosome. *Microbiol. Rev.* **54:** 502–539.
20. 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.
21. Legendre, P. and L. Legendre. 1998. *Numerical Ecology*. Elsevier Science, Amsterdam, The Netherlands.
22. Lupski, J. R. and G. M. Weinstock. 1992. Short, interspersed repetitive DNA sequences in prokaryotic genomes. *J. Bacteriol.* **174:** 4525–4529.
23. Martin, B., O. Humbert, M. Camara, E. Guenzi, J. Walker, T. Mitchell, P. Andrew, M. Prudhomme, G. Alloing, and R. Hakenbeck. 1992. A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. *Nucleic Acids Res.* **20:** 3479–3483.
24. Milkman, R. 1973. Electrophoretic variation in *Escherichia coli* from natural sources. *Science* **182:** 1024–1026.
25. Minana-Galbis, D., M. Farfan, M. C. Fuste, and J. G. Loren. 2004. Genetic diversity and population structure of *Aeromonas hydrophila, Aer. bestiarum, Aer. salmonicida* and *Aer. popoffii* by multilocus enzyme electrophoresis (MLEE). *Environ. Microbiol.* **6:** 198–208.
26. Mora, A., M. Blanco, J. E. Blanco, M. P. Alonso, G. Dhabi, F. Thomson-Carter, M. A. Usera, R. Bartolome, G. Prats, and J. Blanco. 2004. Phage types and genotypes of shiga toxin-producing *Escherichia coli* O157:H7 isolates from humans and animals in Spain: Identification and characterization of two predominating phage types (PT2 and PT8). *J. Clin. Microbiol.* **42:** 4007–4015.
27. Ochman, H. and R. K. Selander. 1984. Standard reference strains of *Escherichia coli* from natural populations. *J. Bacteriol.* **157:** 690–693.
28. Rademaker, J. L., B. Hoste, F. J. Louws, K. Kersters, J. Swings, L. Vauterin, P. Vauterin, and F. J. de Bruijn. 2000. Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system. *Int. J. Syst. Evol. Microbiol.* **50:** 665–677.
29. Rademaker, J. L. W., F. J. Louws, and F. J. de Bruijn. 1998. Characterization of the diversity of ecologically important microbes by rep-PCR fingerprinting, pp. 1–26. *In* A. D. L. Akkermans, J. D. van Elsas, and F. J. de Bruijn (eds.), *Molecular Microbial Ecology Manual*, suppl. 3. Kluwer Academic Publishers, Dordrecht, The Netherlands.
30. Regua-Mangia, A. H., B. C. Guth, J. R. da Costa Andrade, K. Irino, A. B. Pacheco, L. C. Ferreira, V. Zahner, and L. M. Teixeira. 2004. Genotypic and phenotypic characterization of enterotoxigenic *Escherichia coli* (ETEC) strains isolated in Rio de Janeiro city, Brazil. *FEMS Immunol. Med. Microbiol.* **40:** 155–162.
31. 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.
32. 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.

33. Sharples, G. J. and R. G. Lloyd. 1990. A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. *Nucleic Acids Res.* **18:** 6503–6508.
34. Stackebrandt, E., W. Frederiksen, G. M. Garrity, P. A. Grimont, P. Kampfer, M. C. Maiden, X. Nesme, R. Rossello-Mora, J. Swings, H. G. Truper, L. Vauterin, A. C. Ward, and W. B. Whitman. 2002. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* **52:** 1043–1047.
35. Valsecchi, E. 1998. Tissue boiling: A short-cut in DNA extraction for large-scale population screenings. *Mol. Ecol.* **7:** 1243–1245.
36. Vernozy-Rozand, C., M. P. Montet, Y. Bertin, F. Trably, J. P. Girardeau, C. Martin, V. Livrelli, and L. Beutin. 2004. Serotyping, stx2 subtyping, and characterization of the locus of enterocyte effacement island of shiga toxin-producing *Escherichia coli* and *E. coli* O157:H7 strains isolated from the environment in France. *Appl. Environ. Microbiol.* **70:** 2556–2559.
37. Welinder-Olsson, C., M. Badenfors, T. Cheasty, E. Kjellin, and B. Kaijser. 2002. Genetic profiling of enterohemorrhagic *Escherichia coli* strains in relation to clonality and clinical signs of infection. *J. Clin. Microbiol.* **40:** 959–964.
38. Whittam, T. S., M. L. Wolfe, I. K. Wachsmuth, F. Orskov, I. Orskov, and R. A. Wilson. 1993. Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. *Infect. Immun.* **61:** 1619–1629.