5930–5933 Nucleic Acids Research, 1993, Vol. 21, No. 25

RAPD (arbitrary primer) PCR is more sensitive than multilocus enzyme electrophoresis for distinguishing related bacterial strains

Gan Wang[1,2], Thomas S.Whittam[3], Claire M.Berg[2] and Douglas E.Berg[1]*
[1]Departments of Molecular Microbiology and of Genetics, Box 8230, Washington University Medical School, St Louis, MO 63110, [2]Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269 and [3]Institute of Molecular Evolutionary Genetics, Department of Biology, Pennsylvania State University, University Park, PA 16802, USA

Received September 3, 1993; Revised and Accepted November 19, 1993

ABSTRACT

The RAPD (random amplified polymorphic DNA) fingerprinting method, which utilizes low stringency PCR amplification with single primers of arbitrary sequence to generate strain-specific arrays of anonymous DNA fragments, was calibrated relative to the widely used, protein-based multilocus enzyme electrophoretic (MLEE) typing method. RAPD fingerprinting was carried out on five isolates from each of 15 major groups of *Escherichia coli* strains that cause diarrheal disease worldwide (75 isolates in all). Each group consisted of isolates that were not distinguishable from one another by MLEE typing using 20 diagnostic enzyme markers. In our RAPD tests, three or more distinct subgroups in each MLEE group were distinguished with each of five primers, and 74 of the 75 isolates were distinguished when data obtained with five primers were combined. Thus, RAPD typing is far more sensitive than MLEE typing for discriminating among related strains of a species. Despite their different sensitivities, the same general relationships among strains were inferred from MLEE and RAPD data. Thus, our results recommend use of the RAPD method for studies of bacterial population genetic structure and evolution, as well as for epidemiology.

INTRODUCTION

Methods for distinguishing individual bacterial strains and estimating nucleotide sequence diversity are important for detecting and tracing disease outbreaks, and for understanding the genetic structure and evolution of microbial populations. Arbitrary primer PCR (also called 'RAPD' for random amplified polymorphic DNA) (1,2) is one of the most promising of these methods. It uses single oligonucleotides of arbitrarily chosen sequence to prime DNA synthesis at low stringency from pairs of sites to which the oligonucleotide is matched or almost matched. This generates strain-specific arrays of amplified DNA fragments. The formation of these arrays does not depend on prior knowledge of the nucleotide sequence, nor is it affected by DNA modifications that complicate typing by restriction enconuclease digestion of genomic DNA. Because the RAPD method is PCR-based, only nanogram quantities of DNA are required and the DNA need not be double-stranded, highly purified, or of high molecular weight. These features make the RAPD method especially useful for organisms that grow slowly, or for which culturing large volumes is costly or hazardous.

Much of population genetic research over the last two decades has relied on the protein-based, multilocus enzyme electrophoresis (MLEE) method (see 3,4). MLEE detects different alleles of diagnostic genes by scoring the electrophoretic mobilities of the enzymes they encode. Mobility differences usually reflect differences in charge (amino acid sequence) of proteins, and thus a large fraction of single base pair differences in the genes encoding them. In using MLEE data, it is generally assumed that isolates exhibiting the same mobility for a given enzyme contain the same allele of the underlying gene. MLEE studies have shown that relatively few combinations of alleles or multilocus genotypes predominate in some bacterial species. Isolates with the same multilocus genotype have been referred to as 'clones', because of their homogeneity with respect to many variable characters, and an assumption that this similarity reflects recent descent from a common ancestor (3,4).

We present a RAPD analysis of *E.coli* isolates of known MLEE type that helps calibrate the RAPD and MLEE methods. The 75 strains used consist of five isolates from each of 15 MLEE groups that collectively represent 70% of five major O-serogroups of *E.coli* associated with diarrheal disease around the world (Fig. 1; ref. 5). The isolates in any given MLEE group had not been distinguishable from one another by typing with any of 20 diagnostic enzymes.

Our results show general agreement between the grouping of *E.coli* strains obtained by RAPD and by MLEE typing. A recent study of 24 *Trypanosoma cruzi* isolates and several other species of parasitic protozoa also noted a general agreement between

*To whom correspondence should be addressed

MLEE and RAPD typing (7). However, our experiments also show that RAPD typing with a few arbitrary primers is much more sensitive than MLEE typing with 20 diagnostic enzymes, and thus recommends it for epidemiologic and population genetic analyses of diverse organisms.

MATERIALS AND METHODS

Bacteria

Seventy-five *E.coli* clinical isolates belonging to 15 major MLEE-defined groups (five isolates per group) were used (Fig. 1; ref. 5).

DNA preparation

Genomic DNA was extracted from 1.5 ml of stationary phase *E.coli* cells as follows: The cells were spun for 2 min in a microcentrifuge, suspended in 567 μl of 50 mM Tris, 50 mM EDTA, pH 8.0 plus 30 μl 10% SDS and 3 μl of 20 mg/ml proteinase K and incubated 1 hr at 37°C. Then 80 μl of 10% CTAB in 0.7% NaCl was added and the mixture was incubated for 10 min at 65°C. The solution was extracted with 750 μl of chloroform/isoamylalcohol (24:1), spun, and the aqueous phase was re-extracted with phenol/chloroform/isoamylalcohol (25:24:1). DNA was precipitated from the aqueous phase with 500 μl of isopropanol, the precipitate was washed with 70% ethanol, dried briefly and resuspended in 100 μl of 50 mM Tris, 50 mM EDTA, pH 8.0. An aliquot was electrophoresed in a 1% agarose gel and stained with ethidium bromide to estimate the DNA yield, and verify DNA integrity.

RAPD fingerprinting

PCR was carried out in 25 μl containing 20 ng of *E.coli* genomic DNA, 3 mM $MgCl_2$, 20 pmoles of primer, 1 U of AmpliTaq DNA polymerase (Perkin-Elmer Cetus), 250 μM each of dCTP, dGTP, dATP and dTTP (Boehringer) in 10 mM Tris−HCl, pH 8.3, 50 mM KCl and 0.001% gelatin, under a drop of mineral oil. A Perkin-Elmer TC480 thermal cycler was used for amplification. With 10-nt primers, all data used in the present analysis were generated in a cycling program of 45 cycles of [94°C, 1 min; 38°C, 5 min; 38°C ramp to 72°C 3 min, and 72°C, 2 min]. Use of shorter cycling programs (as in refs 2,8) resulted in similar and highly reproducible profiles of bands that differed somewhat from those obtained with this program. With longer (20- or 27-nt) primers, the cycling program was four cycles of [94°C, 5 min; 40°C, 5 min; and 72°C, 5 min], 30 cycles of [94°C, 1 min; 55°C, 1 min; and 72°C, 2 min], plus a final step of 72°C for 10 min (refs 1,8). After PCR, 10−20 μl aliquots of the products were electrophoresed in 2% agarose gels containing 0.5 μg/ml ethidium-bromide in the gel and 1×Tris−acetate running buffer, and the gels were photographed under UV light. The 1 kb DNA ladder (Gibco, BRL) was used as a size marker in all gels.

RESULTS

Ten primers that had been useful in RAPD studies of other microbes (8,9) were screened with a few representative isolates from our collection of diarrheagenic *E.coli* (DEC) strains, and five primers that yielded the largest number of clear bands were selected for detailed testing of sequence divergence among isolates. First, RAPD profiles from one representative of each of the 15 MLEE groups were generated with each primer (Fig. 2). We found that: (i) multiple bands were obtained from each isolate (3 to ~20, depending on the primer and strain); (ii) isolates in closely related MLEE groups yielded similar arrays of RAPD fragments (e.g. from DEC 1 and 2, or from DEC 3, 4 and 5; see Fig. 1); and (iii) each isolate was distinguishable from the others. These results suggest that RAPD tests with five primers are at least as sensitive as MLEE tests with 20 enzymes.

To assess diversity within MLEE groups, four additional isolates from each group (75 isolates, total) were studied with each primer used in Fig. 2. The patterns obtained with a given primer from isolates within any single MLEE group were similar

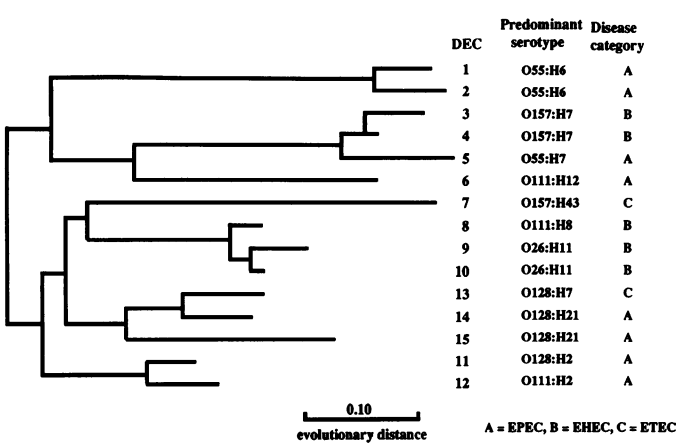

Figure 1. Phylogenetic tree of 15 major groups of diarrheagenic *E.coli* (DEC) isolates determined by MLEE typing (5) and used in this work. The members of each group were homogeneous by MLEE typing with 20 diagnostic enzymes (adapted from 5). Unit of genetic distance = electrophoretically detectable codon changes per locus (6).

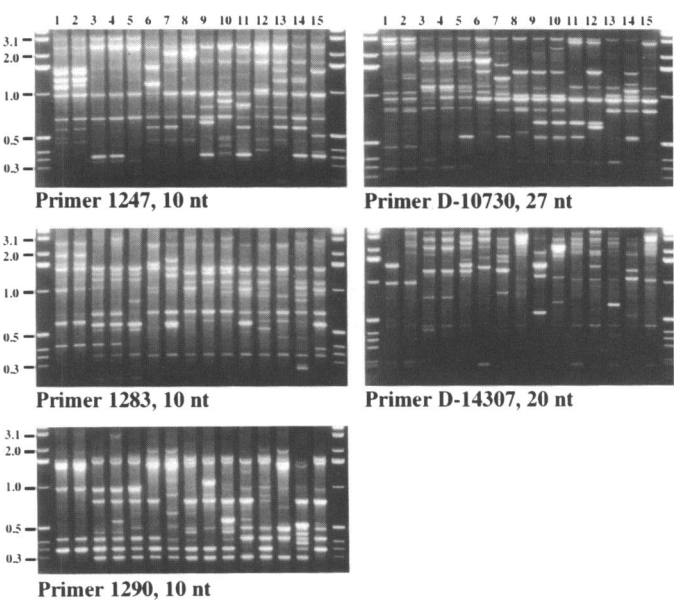

Figure 2. RAPD fingerprinting detects differences among MLEE groups. DNA from one isolate from each of the 15 MLEE groups diagrammed in Figure 1 was used in RAPD amplification with each of five arbitrary primers. The leftmost lane in each panel contains the '1 kb ladder' (Life Technologies Incorporated; Gibco BRL) as a size marker; sizes (in kb) of key fragments are listed to the left.

to each other. For example, with primer D-14307, a 0.95 kb band was obtained from each of the ten isolates, and a 1.45 kb band was obtained from nine of the ten isolates, in the closely related groups DEC 1 and 2 (Fig. 3). These bands were not obtained from isolates in any of the other MLEE groups (the other groups are not closely related to DEC 1 or 2; see Fig. 1). The RAPD tests showed that the isolates in a given DEC group were generally not identical to one another, however. For example, again with primer D-14307, a 0.82 kb and a 0.33 kb band were each generated from just one isolate from the DEC 1 and 2 groups, and a 1.3 kb band was generated from just one of five DEC 1 isolates and two of five DEC2 isolates.

The reproducibility of differences in RAPD profiles was tested using DNAs prepared from different single colony isolates of the same strain on different days. The arrays of RAPD products from these sibling DNA preparations were highly reproducible for fragments up to about 3.5 kb in size for each of 40 strains tested (see Fig. 4). Even the relative intensity of individual bands was highly reproducible (see, for example, faint vs. strong 1 kb bands from DEC 12 strains in Fig. 4).

The most informative of the RAPD primers (D-14307) distinguished each of the five isolates in 11 of the 15 MLEE groups, and 66 different strain types overall (Fig. 3). Each of the other four primers also distinguished at least several distinct strains in each group of five, and 53 or more strains in the set of 75 isolates tested (Table 1). Consideration of the results obtained with each of the five primers distinguished 74 of the 75 strains tested.

The extent of divergence between MLEE groups was estimated using the fraction of bands shared between groups (data available from TSW on request), and compared with the extent of

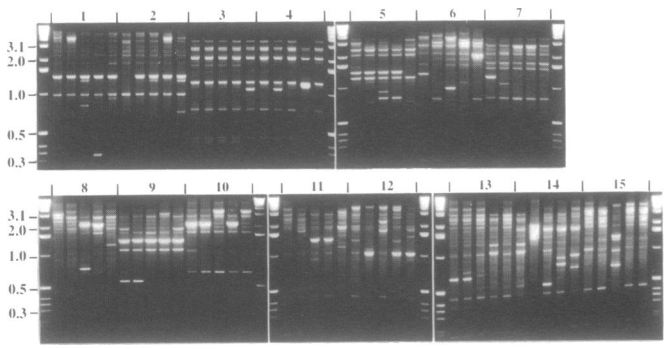

Figure 3. RAPD fingerprinting detects diversity among strains that were not distinguished by MLEE typing. RAPD fingerprinting was carried out with arbitrary primer D-14307 and DNAs from five members of each of the 15 MLEE groups diagrammed in Figure 1. The numbers above each set of five lanes indicate DEC 1–15. The first and last lanes of each panel contain the 1 kb ladder size standard, with key fragment sizes indicated in kb. Nearly as much diversity among isolates in individual groups was seen within and among the DEC groups with the other four primers (Table 1).

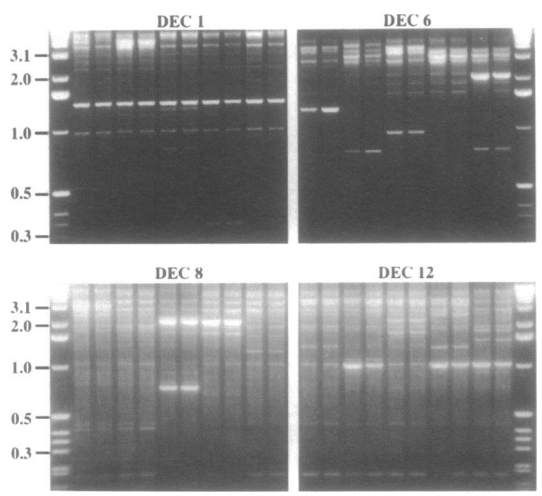

Figure 4. RAPD profiles are reproducible. DNAs from each of two single colony isolates from each of 20 isolates were used in parallel in RAPD tests with primer D-14307. The first or last lanes of each panel contain the 1 kb ladder size standard, with key fragment sizes indicated in kb.

Table 1. RAPD patterns distinguish isolates within MLEE groups

Primer	1247	1283	1290	10730	14307
DEC group	Number of types distinguished				
1	1	3	1	5	5
2	2	3	2	4	5
3	1	2	3	3	2
4	1	4	5	1	4
5	4	4	5	5	5
6	4	5	3	2	5
7	4	2	3	4	5
8	5	5	5	4	5
9	5	4	5	3	5
10	5	3	4	5	4
11	5	4	5	3	5
12	4	5	3	4	5
13	3	4	4	4	5
14	5	4	3	4	5
15	2	3	3	4	2
No. bands	34	40	30	44	41
No. types	53	54	53	63	66

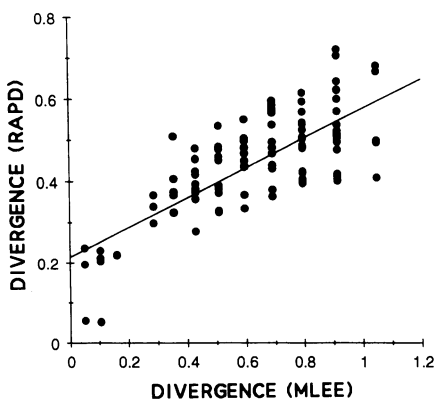

Figure 5. Comparision of divergence calculated from MLEE and RAPD test results. Divergence between DEC clones, based on MLEE data (number of protein electrophoretic differences detected per locus), was estimated as standard genetic distance, $d = -\ln(I)$ where I is the proportion of 20 enzyme loci with matching alleles (6). Divergence between DEC clones, based on RAPD data, was measured as 1 – the fraction of shared RAPD bands. One hundred and five pairwise comparisons were carried out. There was a significant correlation between the two measures of divergence (R=0.76).

divergence estimatd from MLEE data (Fig. 5). The correlation coefficient was 0.76 for the 105 pairwise comparisons, indicating close agreement between estimates of divergence based on MLEE and RAPD typing data.

DISCUSSION

Our results showed that RAPD tests with just one or a few short arbitrary primers are more sensitive than conventional multilocus enzyme electrophoresis (MLEE) for detecting differences between closely related *E.coli* isolates. We studied 75 isolates drawn from 15 major MLEE groups of diarrheagenic *E.coli* strains from around the world. While MLEE tests with 20 enzymes separated the 75 isolates into 15 groups, RAPD tests with just five primers distinguished 74 of the 75 isolates, but showed greater inter-group than intra-group variation. It is thus clear that individual MLEE groups of diarrheagenic *E.coli* strains are quite diverse, and that RAPD typing has far greater power to discriminate among them than does MLEE typing.

Consistent with the diversity seen here, MLEE tests with 19 additional enzymes subdivided some of the original 15 groups, and thus distinguished 25 groups overall in this collection of 75 strains (10). *E.coli* strains have also been distinguished by pulsed field gel electrophoresis (11), Southern blotting with insertion sequence (IS) probes (12), and phage typing (13). However, each of these methods is less efficient than RAPD typing when numerous isolates are to be examined. Given the high sensitivity and efficiency of RAPD typing, it is noteworthy that the relationships among strains inferred from its use are in good agreement with those inferred from MLEE typing (Fig. 5).

Statistical methods for relating RAPD data to sequence divergence have been developed recently (14), by assuming that amplification depends on an exact match in sequence between the oligonucleotide primer and the primer binding site, and that the disappearance (or appearance) of a band in a profile reflects nucleotide sequence substitutions in the primer binding site. In recent Southern hybridization tests using cloned RAPD fragments from *E.coli* K-12 laboratory strains, we identified two fragments for which polymorphisms can be attributed to point mutations, but also two others for which polymorphisms often reflect deletion events (15). Corrections to allow more accurate estimates of rates of nucleotide sequence divergence within and between MLEE groups, and thereby a better understanding of forces guiding microbial evolution, could be obtained by further characterization of additional polymorphic RAPD bands. Even without these calibrations, however, it is clear that RAPD tests provide a sensitive, efficient and reliable means of distinguishing closely related strains, and are thus of great value in epidemiologic and evolutionary studies.

ACKNOWLEDGEMENTS

This collaboration grew out of discussions supported by a Foundation for Microbiology (American Society for Microbiology) lectureship to DEB. We are grateful to S.Ake and N.Akopyanz for stimulating discussion, and to J.Woods for critical reading of the manuscript. This research was supported by grant HG00563 to DEB, and AI24566 and AI00964 to TW from the US Public Health Service and DE-FGO2-90ER-610 to CMB from the Department of Energy.

REFERENCES

1. Welsh, J. and McClelland, M. (1990) Nuc. Acids Res. **18**,7213–7218.
2. Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. (1990) Nuc. Acids Res. **18**,6531–6535.
3. Selander, R. K. and Musser, J. (1990). In B. H. Iglewski and V. L. Clark (eds.) Molecular Basis of Bacterial Pathogenesis (The Bacteria, Vol. 11) Academic Press, Inc., San Diego. pp 11–36.
4. Maynard Smith, J., Smith, N.H., O'Rourke, M. and Spratt, B.G. (1993) Proc. Natl. Acad. Sci. USA **90**,4384–4388.
5. Whittam, T.S., Wolfe, M.L., Wachsmuth, I.K., Ørskov, F, Ørskov, I. and Wilson, R.A, (1993) Infect. Immun. **61**,1619–1629.
6. Nei, M. (1987) Molecular Evolutionary Genetics. Columbia University Press, NY.
7. Tibayrenc, M., Neubauer, K., Barnabe, C., Guerrini, F. Skarecky, D. and Ayala, F.J. (1993) Proc. Natl. Acad. Sci. USA **90**,1335–1339.
8. Akopyanz, N., Bukanov, N. Westblum, T.U., Kresovich, S. and Berg, D.E. 1992. Nuc. Acids. Res. **20**,5137–5142.
9. Kersulyte, D., Woods, J.P., Keath, E.J., Goldman, W.E., and Berg, D.E. (1992) J. Bacteriol. **174**,7075–7079.
10. Miller, B.J. and Whittam, T.S. unpublished.
11. Arbeit, R.D., Arthur, M., Dunn, R., Kim, C., Selander, R.K. and Goldstein, R. (1990). J. Infect. Dis. **161**,220–235.
12. Lawrence, J.G., Dykhuizen, D.E., DuBose R.F. and Hartl, D.L. (1989) Molec. Biol. Evol. **6**,1–14.
13. Frost, J.A., Cheasty, T., Thomas, A. and Rowe, B. (1993) Epidemiol. Infect. **110**,469–475.
14. Clark, A.G. and Lanigan, C.M.S. (1993) Molec. Biol. Evol. **10**,1096–1111.
15. Brikun, I., Suziedelis, K. and Berg, D. E. submitted for publication.