JOURNAL OF BACTERIOLOGY, May 1989, p. 2528–2532
0021-9193/89/052528-05$02.00/0

Vol. 171, No. 5

Hypervariable DNA Fingerprinting in *Escherichia coli*: Minisatellite Probe from Bacteriophage M13

BING HUEY AND JEFF HALL*

School of Public Health, University of California at Berkeley, Berkeley, California 94720

Received 30 September 1988/Accepted 21 February 1989

Extensive restriction-fragment-length polymorphism was revealed in *Escherichia coli* strains by using a region of the bacteriophage M13 genome as a DNA hybridization probe. This variation was observed across natural strains, in clinical samples, and to a lesser extent in laboratory strains. The sequence in M13 which revealed this fingerprint pattern was a region of the gene III coat protein, which contains two clusters of a 15-base-pair repeat. Oligonucleotides made to a consensus of these repeats also revealed the fingerprint profile. While this consensus sequence has significant homology to the lambda chi site sequence, an oligonucleotide made of the chi sequence did not reveal polymorphic fingerprint patterns in *E. coli*. The strain variation revealed by the M13 and M13-derived oligonucleotide probes will be useful for bacterial characterization and should find use in studies of bacterial evolution and population dynamics. The findings raise questions about what these repeated sequences are and why they are so variable.

Knowledge about genetic variation within a bacterial species is useful for identifying and typing pathogenic strains as well as for examining evolutionary divergence and population dynamics. Various methods exist for observing genetic variation in bacterial strains, including those used to determine enzyme profiles, differences in metabolic pathways, and drug resistance phenotypes (7). However, it is often desirable to subtype even more precisely than these methods allow. At the extreme, DNA sequence analysis can reveal changes in a single base pair; however, sequence analysis is laborious and requires knowledge of the exact sequence within which such differences have occurred.

Recently, the application of DNA hybridization probes to moderately repeated hypervariable minisatellite sequences in humans and other eucaryotes allowed researchers to genetically fingerprint an organism (5, 16). This approach has the precision required for paternity testing (8) and forensics (3) as well as for providing useful genetic information for the construction of genetic maps involving large pedigrees (17). Such minisatellite probes have been obtained serendipitously by a number of laboratories, usually in the course of investigating tandemly repeated sequences observed in cloned DNA fragments (5, 10, 16). In the genomes of humans and other eucaryotes such tandem repeats are found in multiple copies, often exhibit length variation in a given population, and have been shown to be inherited in a Mendelian manner (4). Some researchers have hypothesized that these sequences could be sites for recombination, which would account for both their abundance and their variability (5).

While cloning a specific human sequence related to a minisatellite probe derived from the genome of bacteriophage M13, we observed related sequences in *Escherichia coli*. Examination of a number of strains of *E. coli*, including laboratory strains, clinical isolates, and strains from natural populations, revealed the presence of these sequences in all strains examined and showed that they were highly variable among strains. This probe should prove useful for typing strain differences and for quantifying evolutionary divergence. Further study of these sequences should reveal many biological features accounting for their ubiquitous nature, variability, and function in *E. coli*.

MATERIALS AND METHODS

Bacterial strains. The eight laboratory strains of *E. coli* (LE392, BB4, DP50, NM538, XL1, HB101, JM103, and NM539) were in use in our laboratory at the time for cloning purposes or were obtained from other laboratories. The 72 natural strains were obtained from Ochman (11). All strains were grown on YT medium (8 g of tryptone, 5 g of NaCl, and 5 g of yeast extract per liter) at 37°C. DNA from the clinical isolates was kindly provided by Jeff Lawrence of Washington University, St. Louis, Missouri.

DNA isolation. Fifteen milliliters of a culture grown to confluence overnight was centrifuged at 3,000 rpm for 20 min in a Sorvall HS4 rotor. The bacterial pellet was suspended in 10 ml of 10 mM Tris base–1 mM EDTA (pH 7.5) and incubated with lysozyme for 10 min at 4°C to a final concentration of 4 mg/ml. Four milliliters of 0.5 M EDTA was added, and incubation was continued at 4°C for an additional 20 min. Proteinase K was added to a final concentration of 0.2 mg/ml, sodium dodecyl sulfate was added to a final concentration of 0.5%, and the solution was incubated overnight at 55°C with shaking. The lysate was extracted with an equal volume of phenol and then with an equal volume of chloroform and dialyzed overnight against 10 mM Tris base–1 mM EDTA (pH 7.5) at 4°C. The DNA was precipitated by adding 1/10 volume of 3 M sodium acetate and 2 volumes of ethanol, incubated at −20°C overnight, centrifuged at 4,000 × *g* for 20 min, and suspended in 10 ml of 10 mM Tris base–1 mM EDTA (pH 7.5). The DNA was incubated at 37°C for 2 h with RNase at a concentration of 1 mg/ml, followed by the addition of sodium dodecyl sulfate to 0.5% and 1 mg of proteinase K. This was incubated overnight at 55°C with shaking. The DNA was extracted with an equal volume of phenol and then with an equal volume of chloroform and dialyzed overnight against 10 mM Tris base–1 mM EDTA (pH 7.5) at 4°C.

* Corresponding author.

Southern hybridization. DNA (15 μg) was incubated overnight with three times the units of restriction enzyme recommended by the manufacturer for complete digestion. Then a second portion of three times the required enzyme units was incubated for an additional 5 h to ensure complete digestion. Restriction enzymes were purchased from New England BioLabs, Inc., Beverly, Mass. Digested DNA was electrophoresed in 0.8% agarose at 2 V/cm for 18 h, and DNA fragments were transferred to nitrocellulose by the sandwich blot method of Southern (14).

Filters were hybridized with DNA probes prepared as follows. The + strand of M13 (obtained from Bethesda Research Laboratories, Inc., Gaithersburg, Md.) was labeled by the random priming method of Feinberg and Vogelstein (2) using the Amersham Multiprime labeling system (Amersham Corp., Arlington Heights, Ill.). Oligonucleotides were purchased from Synthetic Genetics, San Diego, Calif., and were constructed with the 6-base recognition sequence of *Bam*HI at the 3′ termini. The oligonucleotides were labeled by using *Bam*HI linkers (purchased from New England BioLabs) as the initiation primers in the Klenow-mediated extension reaction. All probes were labeled to a specific activity of 10^8 cpm/mg with [α-^{32}P]dCTP purchased from Amersham. The concentration of probe DNA in the hybridization mixture was 10 ng/ml.

The hybridization mixture consisted of 20% formamide, 0.1% sodium dodecyl sulfate, 5× SSPE (from 20× stock; 20× SSPE is 7.4 g of disodium EDTA, 6.4 g of NaOH, 24 g of NaH_2PO_4, and 210 g of NaCl per liter), 5× Denhardt solution (from 100× stock; 100× Denhardt solution is 20 g of Ficoll, 20 g of polyvinylpyrrolidone, and 20 g of bovine serum albumin per liter), and 0.5% powdered milk. This preparation was also used for prehybridization. Prehybridization was done at 42°C for 24 h; hybridization was also done at 42°C for 24 h but with agitation. Filters were rinsed in 2× SSPE–0.1% sodium dodecyl sulfate for 15 min at room temperature and exposed overnight at −70°C with an intensifying screen.

RESULTS

Hybridization of M13 and an M13-derived oligonucleotide to laboratory *E. coli* strains. This study was begun during initial attempts to clone an M13-related sequence from the human genome. It has been observed that DNA probes derived from the regions of M13 from bases 1833 to 1894 and from bases 2283 to 2401 detect hypervariable minisatellite sequences in humans and other eucaryotes (16). These regions are part of the M13 gene III coat protein and contain 4 and 9 copies, respectively, of a 15-base-pair repeat. During our initial screening of a human genomic lambda library propagated on *E. coli* strain LE392, extremely high background from the host strain led us to determine whether this strain had sequences related to the repeat sequence from M13. We examined seven additional common laboratory strains: BB4, DP50, NM538, XL1, HB101, JM103, and NM539. Figure 1A shows the results obtained when probing *Pst*I-digested genomic DNA from these strains with the 30-base oligonucleotide $(GAGGGTGGCGGTTCT)_2$, containing two copies of a consensus sequence of the M13 repeat. All eight strains had multiple copies of a sequence related to the 15-base-pair M13 consensus sequence; however, there were few differences in the hybridization patterns among the strains. The same *Pst*I-digested DNA, when hybridized to the entire radiolabeled M13 molecule, showed multiple copies related to sequences in M13 but exhibited slightly more variability than the oligonucleotide pattern (Fig. 1B). This difference is probably the result of hybridization by variants of the consensus sequence which are present when labeling the entire repeated region, as opposed to the homogeneous sequence of the oligonucleotide. It should be noted that the pattern obtained with radiolabeled probe made from the oligonucleotide had a slightly greater representation of fragments in the 1- to 3-kilobase range, while the entire M13 molecule as a probe revealed more fragments between 3 and 15 kilobases (Fig. 1). This is probably due to the limited probe length (30 base pairs) that can be generated by the oligonucleotide.

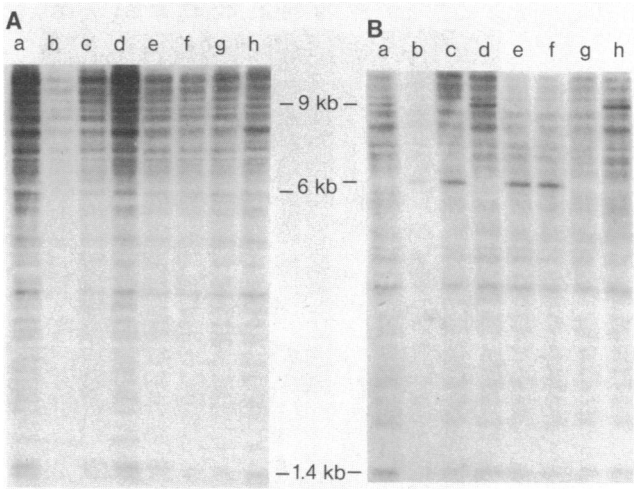

FIG. 1. Hybridization of the 15-base-pair oligonucleotide (A) and the entire M13 molecule (B) to 15 mg of *Pst*I-digested DNA obtained from eight laboratory strains of *E. coli*. The strains used were BB4 (lanes a), DP50 (lanes b), HB101 (lanes c), JM103 (lanes d), LE392 (lanes e), NM538 (lanes f), NM539 (lanes g), and XL1 (lanes h). kb, Kilobases.

To verify that the repeat-containing portion of the M13 molecule is responsible for the fingerprint pattern, M13 replicative-form DNA was digested with *Hae*III, and various fragments were gel purified, radiolabeled, and used as hybridization probes on replicate blots of *Pst*I-digested *E. coli* DNA. Only the *Hae*III fragments derived from the region between bases 1396 to 2554, which contains the two blocks of repeats (1), hybridized to *E. coli* DNA (data not shown). The negative result with the rest of the M13 molecule also shows that the *E. coli* strains examined do not have any endogenous M13 which could account for the fingerprint pattern we observed.

Hybridization of M13 to natural and clinical strains of *E. coli*. The minor variation among the eight laboratory-derived strains detected with the whole M13 molecule and the oligonucleotide as probes prompted us to examine strains of *E. coli* which might be more divergent in origin. Since the eight laboratory strains have all been derived from related populations in the past 50 years, we examined DNA from a large collection of reference strains cataloged by Ochman and Selander (11) from natural *E. coli* populations. These 72 reference strains were isolated from a variety of sources (human, dog, orangutan, giraffe, sheep, etc.) and from various geographical locations (the United States, Europe, New Guinea, Bali, etc.). The strains have been characterized by multilocus enzyme profiles and clearly represent a large range of genotypic variation.

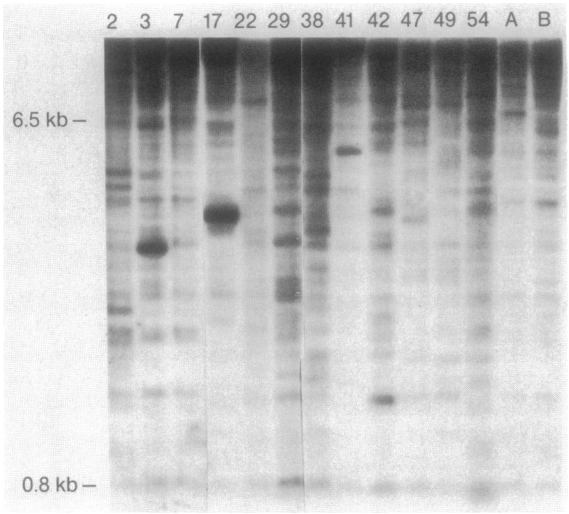

FIG. 2. Hybridization of the M13 molecule to 15 mg of *Pst*I-digested DNA isolated from natural strains of *E. coli*. The first 12 lanes represent natural strains from the catalog of Ochman and Selander (11), and the last 2 lanes represent two laboratory strains, LE392 (A) and XL1 (B). kb, Kilobases.

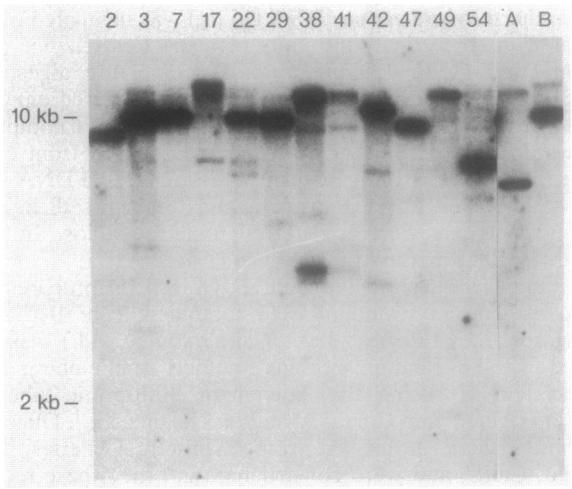

FIG. 3. Hybridization of the M13 molecule to 15 mg of *Pst*I-digested DNA from natural *E. coli* strains at increased hybridization stringency. DNA from the strains represented in Fig. 2 was hybridized at 30% formamide instead of 20%; otherwise all conditions were identical to those used for the experiment represented in Fig. 2. Strain numbers are listed above each lane, with the last two lanes representing laboratory strains LE392 (A) and XL1 (B). kb, Kilobases.

Figure 2 shows *Pst*I-digested DNA from 12 of the natural strains and 2 of the laboratory strains probed with radiolabeled M13. In contrast to the laboratory strains represented in Fig. 1, the natural strains showed extensive variation in hybridization patterns. These differences allow each strain to be distinguished on the basis of its M13 hybridization profile. Under these hybridization conditions more than 20 bands were present in each lane (Fig. 2), ranging in size from less than 1 to more than 20 kilobases. Pairwise comparison allowed all strains to be classified relative to each other. As an index of the variation detected by the M13 fingerprint pattern, differences based on the presence or absence of bands were scored. An example of such an analysis for 13 of the 72 natural *E. coli* strains is shown in Table 1. In these natural strains, differences between pairs ranged from 7 to 27 bands.

When the natural strains were probed with radiolabeled M13 under more stringent conditions (30% instead of 20% formamide), most of the fainter bands were lost (Fig. 3). In almost all the strains at least one very intense band and several fainter bands remained. All the strains could again be distinguished on the basis of their hybridization profiles, but under these conditions there was much less information present for the determination of differences.

We also analyzed a number of *E. coli* strains obtained from bacteremic patients at a large hospital. The amount of variation among these strains was between that observed in the natural strains and that seen in the laboratory strains (Fig. 4). All 12 strains represented in Fig. 4 are distinguishable by the presence or absence of at least one band in any pair. For example, by using the eight bands marked A through H in Fig. 4, the 12 strains can all be distinguished from one another. In the clinical strains, the number of bands that differed between pairs ranged from 2 to 9, compared with the 7- to 27-band difference seen in the natural strains. These clinical isolates all had identical electromorphic mobility profiles, as determined with 12 enzymes encoded at separate chromosomal locations on the *E. coli* genome (12). Thus, the ability to distinguish them on the basis of a simple DNA hybridization test would prove useful in a clinical setting.

TABLE 1. Analysis of variation among 13 natural *E. coli* strains in patterns observed when probing *Pst*I-digested DNA with the M13 molecule[a]

Strain no.	No. of differences between pairs for strain no.:												
	2	3	7	17	22	29	38	41	42	47	49	54	58
2		18	15	24	12	25	25	20	22	19	25	31	27
3			15	23	22	27	17	21	21	25	15	22	22
7				20	15	22	18	24	18	15	17	23	14
17					13	22	19	22	16	21	18	16	22
22						18	18	24	20	17	12	18	15
29							20	17	14	19	11	15	17
38								7	14	16	9	17	19
41									22	19	16	19	18
42										21	22	20	19
47											20	18	20
49												16	13
54													0
58													

[a] Pairwise comparisons were made for all possible pairs among the 13 strains (collected by Ochman and Selander [11]). A difference was scored each time a band appeared in one strain that did not appear in the other of the pair. The total number of differences noted for each pairwise comparison appears in the appropriate cell of the table.

DISCUSSION

Use of the tandem repeat sequence found in the coat protein gene of bacteriophage M13 as a hybridization probe in *E. coli* allows the detection of considerable strain variation in both natural and clinical strains. Detection of this variation provides a rapid means for confirming bacterial strain identity and should prove useful for observing evolutionary divergence in clinical and natural bacterial populations. Our pairwise comparisons showed the laboratory strains differing by 0 to 1 bands, the clinical strains differing

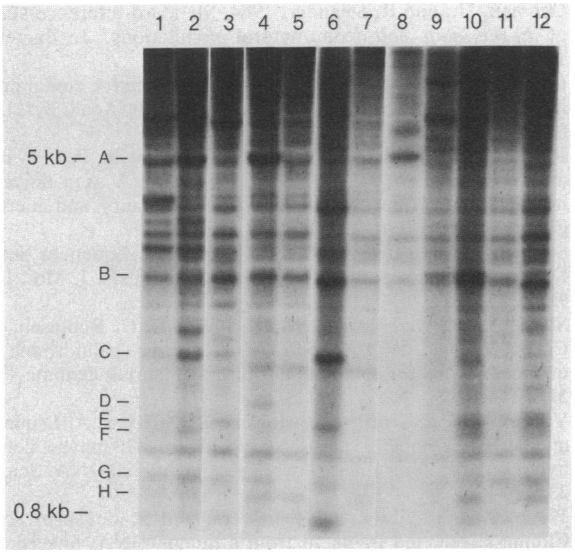

FIG. 4. Hybridization of the M13 molecule to 15 mg of *Pst*I-digested DNA from 12 clinical strains of *E. coli*. Bands by which the strains can be distinguished from one another are labeled A to H. kb, Kilobases.

by 2 to 9 bands, and the natural strains differing by 7 to 27 bands. Thus, closely related *E. coli* strains share more M13-related bands and show fewer differences in a pairwise comparison like that in Table 1.

These findings raise questions as to what the specific sequence or sequences detected are and what biological functions they have. We suspect that the fingerprint patterns generated by M13 are the result of sequences scattered throughout the *E. coli* genome which exhibit length variation due to variable numbers of a tandem repeat which has homology to the 15-base-pair M13 repeat unit. The polymorphisms are probably due to length variation because, along with *Pst*I, we have detected the same variation with the restriction enzymes *Hin*fI, *Ava*II, *Bgl*II, *Bam*HI, and *Hin*dIII (data not shown).

It may be asked why no one else has observed the fingerprint patterns described here. We believe that the success of our approach is due in part to the hybridization conditions—both the low-formamide and low-stringency washes after hybridization—and, more importantly, to the use of hybridization and prehybridization cocktails containing milk proteins instead of sheared DNA as the nonspecific blocking reagent. Hybridization cocktails containing DNA will block any signal of a ubiquitous, repetitive nature in the probe DNA and cannot be used to detect these patterns (6).

We conducted a search of sequences from *E. coli* in the GENBANK data base for homology with the M13 tandem repeat. Sequences were found which exhibited limited homology, including those in the *birA* gene, which had 11 direct matches with the 15-base-pair consensus sequence, and the *rgnB* gene, which had 10 direct matches. The regions of homology found in our search occurred most often in noncoding regions (both the *birA* and *rgnB* homologies were outside the reading frame); however, several were found within the coding region. Since the M13 repeat is within the coding region of the gene III coat protein, being outside coding regions is not a universal characteristic of these sequences. None of the sequences identified in our search as having homology with the M13 repeat were themselves in obvious repeated arrays.

A large degree of homology with the lambda chi site was observed, as shown below:

```
chi       – G C T G G T G G
                | | | | |
M13 15-mer – G A G G G T G G C G G T T C T
                  | | |   | |
                  G C T G G T G G – chi
```

This seemed very interesting because chi is a known site for recombination in lambda (9), and the large amount of variation observed across the *E. coli* strains could be due to recombination events. An oligonucleotide probe was constructed consisting of three copies of the 8-base-pair chi sequence. However, under our least-stringent hybridization conditions (20% formamide), this oligonucleotide did not detect a polymorphic pattern and showed only a few very faint and constant bands across a number of the natural *E. coli* strains (data not shown). We are therefore certain that the M13 fingerprint pattern is not caused by hybridization to endogenous chi sequences in *E. coli*. The REP sequence described by Stern et al. (15) is thought to be a dispersed, sometimes tandemly repeated sequence; however, we failed to detect homology between REP and the M13 tandem repeat in the GENBANK computer search, and we believe that REP and the M13-related sequences are not related. No homology was detected to any known insertion elements or transposons found in the GENBANK data base, but the sequence which the M13 repeat detects might function in a related manner, which would also account for the large variability seen across strains. The minisatellite probes developed by A. J. Jeffreys, which have been extensively used in humans as well as in dogs, cats, birds, and mice, was also tested in our study for homology to sequences present in the *E. coli* genome. We did not detect any signal when using the Jeffreys 6.3 minisatellite probe in *E. coli* DNA (unpublished results).

We have limited our work to strains of *E. coli*. However, other bacteria may have similar polymorphic sequences, in which case M13 profiles could be of use in classifying these strains. Indeed, *B. subtilis* also contains sequences homologous to the M13 repeat (unpublished results). Sequences homologous to the M13 repeat have been observed in organisms ranging from humans to *E. coli* (13). While the biological function of this ubiquitous sequence remains to be elucidated, it may prove extremely interesting if the widespread distribution of the sequence is an indication of its importance.

ACKNOWLEDGMENTS

We thank Mary-Claire King, in whose laboratory this work was carried out, for her enthusiastic indulgence; H. Ochman, who provided invaluable material and advice; and P. Leon, D. Durnam, G. Chang, G. Ames, and C. Orrego for careful reading of the manuscript.

This work was supported by Public Health Service grant CA27632 from the National Institutes of Health.

LITERATURE CITED

1. **Devor, E. J., A. K. Ivanovich, J. M. Hickok, and R. D. Todd.** 1988. A rapid method for confirming cell line identity: DNA "fingerprinting" with a minisatellite probe from M13 bacteriophage. Biotechniques **6(3)**:200–201.
2. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
3. **Gill, P., A. J. Jeffreys, and D. J. Werrett.** 1985. Forensic

application of DNA 'fingerprints.' Nature (London) **318**:577–579.

4. **Jeffreys, A., V. Wilson, R. Kelley, B. Taylor, and G. Bulfield.** 1987. Mouse DNA fingerprints: analysis of chromosome localization and germ line stability of hypervariable loci in recombinant inbred strains. Nucleic Acids Res. **15**:2823–2836.
5. **Jeffreys, A., V. Wilson, and S. Thein.** 1985. Hypervariable minisatellite regions in human DNA. Nature (London) **314**: 67–73.
6. **Johnson, D., J. Gautsch, J. R. Sportsman, and J. Elder.** 1984. Improved technique utilizing nonfat dry milk for analysis of proteins and nucleic acids transferred to nitrocellulose. Gene Anal. Tech. **1**:3–8.
7. **Lennette, E., A. Balows, W. Hausler, and H. J. Shadomy (eds.).** 1985. Manual of clinical microbiology. American Society for Microbiology, Washington, D.C.
8. **Lewin, R.** 1986. DNA fingerprints in health and disease. Science **233**:521–522.
9. **Malone, R. E., D. K. Chattoraj, D. H. Faulds, M. M. Stahl, and F. W. Stahl.** 1978. Hotspots for generalized recombination in the *E. coli* chromosome. J. Mol. Biol. **121**:473–491.
10. **Nakamura, Y., M. Leppert, P. O'Connell, R. Wolf, T. Holm, M. Culver, C. Martin, E. Fukimoto, M. Hoff, E. Kumlin, and R. White.** 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science **235**:5616–1622.
11. **Ochman, H., and R. Selander.** 1984. Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. **157**:690–693.
12. **Ochman, H., and R. Selander.** 1984. Evidence for clonal population structure in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **81**:198–201.
13. **Ryshov, A. P., A. G. Jincharadze, M. I. Prosnyah, P. L. Ivanov, and S. A. Limborsha.** 1988. M13 phage DNA as a universal marker for DNA fingerprinting of animals, plants, and microorganisms. FEBS Lett. **233**:388–392.
14. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
15. **Stern, M. J., G. F. Ames, N. H. Smith, E. C. Robinson, and C. F. Higgins.** 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. Cell **37**:1015–1026.
16. **Vassart, G., M. Georges, R. Monsieur, H. Brocas, A. Leguarre, and D. Christophe.** 1987. A sequence in M13 phage detects hypervariable minisatellites in human and animal DNA. Science **235**:683–684.
17. **Wong, Z., V. Wilson, A. J. Jeffreys, and S. L. Thein.** 1987. Cloning a selected fragment from a human DNA 'fingerprint': isolation of an extremely polymorphic minisatellite. Nucleic Acids Res. **14**:4605–4616.