Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 8902–8906, November 1989
Genetics

DNA fingerprinting by sampled sequencing

(ordered clone bank/fluorescent reporters/genome mapping/biotin)

SYDNEY BRENNER* AND KENNETH J. LIVAK†‡

*Medical Research Council Molecular Genetics Unit, Hills Road, Cambridge CB2 2QH, England; and †Central Research & Development Department, E. I. du Pont de Nemours & Company, Experimental Station, P.O. Box 80328, Wilmington, DE 19880-0328

Contributed by Sydney Brenner, August 14, 1989

ABSTRACT **We describe a method for characterizing DNA segments that combines limited sequencing with size separation of restriction fragments. As part of a multistep procedure, 5′ overhangs of unknown sequence are generated by cleavage with a class IIS restriction enzyme. After labeling of these ends by using dideoxynucleotides tagged with distinctive fluorescent dyes, the restriction fragments are analyzed by polyacrylamide gel electrophoresis and detection of fluorescent emissions using a commercially available DNA sequencer. The nucleotide-specific fluorescent signatures permit determination of the terminal sequence for each labeled end. The set of labeled fragments, characterized by both size and terminal sequence, constitutes a fingerprint that can be used to compare DNA segments for overlap or relatedness. The inclusion of terminal sequence data dramatically increases the information content of the fingerprint, making comparisons more reliable and efficient than those based upon size alone.**

An important approach to the physical mapping of complex genomes involves ordering adjacent DNA fragments in a DNA clone collection. Methods based on restriction enzyme analysis have been used to order large numbers of cloned DNA segments derived from *Caenorhabditis elegans*, yeast, and *Escherichia coli* (1–4). These methods, based on comparisons of restriction fragment length and/or restriction site order, depend critically on the amount of information that can be extracted from each clone to make a reliable assignment of overlaps between clones. A qualitative view of the amount of information required to match two clones can be obtained in the following way. Consider a bacterium with 4 million (= 4^{11}) base pairs. On the average, an 11-base sequence occurs uniquely in such a genome, and this means that an overlap can be established between two pieces of DNA if they have identical undecanucleotides. An 11-base sequence (4^{11}) is equivalent to 22 bits (2^{22}) of information. By the same token, humans, with 4^{16} base pairs of DNA, have 32-bit genomes. How can equivalent information be obtained from the analysis of restriction fragments?

In the fingerprinting procedure of Coulson *et al.* (1), the bands were separated in sequencing gels which can accurately resolve 256 (= 2^8) oligonucleotides successively differing by one nucleotide in length. It is irrelevant how this pattern is obtained; each band contributes 8 bits of information. Thus, a coincidence of three bands, which provides 24 bits, will be necessary to establish overlap of clones from a bacterial genome. For the human genome, at least four bands would be required to provide the minimum of 32 bits. Notice that increasing the resolution of a gel to fragment lengths of 500 nucleotides adds only one additional bit of information to each band. Agarose gel separations, used in restriction enzyme mapping experiments, would require more bands since they have very much lower resolution, perhaps 5 or 6 bits per band. Low information content per band requires sampling a large fraction of the clone; this, in turn, restricts the subset of a random library for which sufficient information can actually be obtained to establish overlaps. Therefore, many more clones than expected need to be examined to populate this subset (5). Awkward distributions of restriction enzyme sites used to analyze the DNA make matters worse and further restrict the subset of clones that can be overlapped. Thus, the best way to improve this method would be to increase the amount of information obtained from each clone. Here we describe a fluorescent fingerprinting procedure that couples band separation with sampled nucleotide sequencing, adding this information to the length information used previously.

MATERIALS AND METHODS

Direct Fingerprinting Procedure. Primary cleavage was performed by incubating 1 μg of φX174 replicative form DNA (New England Biolabs) with *Fok* I (New England Biolabs) in the supplier's recommended buffer. After precipitation with ethanol from 2.5 M ammonium acetate, the DNA was dissolved in 9 μl of 10 mM Tris·HCl, pH 8.0/1 mM EDTA. Nucleotide-specific labeling of the *Fok* I ends was achieved by adding 4 μl of 5× reverse transcriptase buffer (250 mM Tris·HCl, pH 8.3/150 mM KCl/40 mM $MgCl_2$/5 mM dithiothreitol), 2 μl of a solution containing dNTPs at 5 μM each, 4 μl of a solution containing a mixture of succinylfluorescein-labeled 2′,3′-dideoxynucleoside triphosphates (SF-ddNTPs) (6), and 1 μl (4.5 units) of avian myeloblastosis virus reverse transcriptase (New England Nuclear). The SF-ddNTP solution contained 2.5 μM SF-505-ddGTP, 31.25 μM SF-512-ddATP, 7.5 μM SF-519-ddCTP, and 31.25 μM SF-526-ddTTP (numbers indicate various fluorescein derivatives). The SF-ddNTPs were obtained from New England Nuclear as 125 μM stocks and diluted in 10 mM Tris·HCl, pH 8.0/1 mM EDTA. The labeling reaction mixture was incubated at 42°C for 10 min. The DNA was recovered by precipitation with ethanol, dissolved in 16 μl of 10 mM Tris·HCl, pH 7.4/1 mM EDTA, and incubated at 68°C for 10 min. Secondary cleavage was performed by incubating the sample with *Hin*fI (Pharmacia) in the supplier's recommended buffer. The sample was passed over a spun Sephadex G-50 column (5 Prime → 3 Prime, West Chester, PA) that had been rinsed with water. The eluate from the column was dried under reduced pressure in a Speed-Vac concentrator (Savant Instruments). The residue was suspended in 10 μl of 95% (vol/vol) formamide/12.5 mM EDTA, incubated in a boiling water bath for 5 min, and cooled on ice. One microliter of this sample was loaded on a denaturing polyacrylamide gel consisting of 6% acrylamide/bisacrylamide (19:1, wt/wt); 8 M urea in electrophoresis buffer (10× buffer: 162 g of Tris base, 27.5 g of boric acid, and 9.3 g of $Na_2EDTA{\cdot}2H_2O$ dissolved

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: SF-ddNTPs, succinylfluorescein-labeled 2′,3′-dideoxynucleoside triphosphates.

‡To whom reprint requests should be addressed.

in H_2O to a total volume of 1 liter). The gel was electrophoresed and analyzed for fluorescent emissions by using the Genesis 2000 DNA analysis system (DuPont), following the manufacturer's instructions.

Capture Fingerprinting Procedure. Primary cleavage and attachment of biotin were performed simultaneously by incubating each DNA sample in a 40-μl reaction mixture containing 25 mM Tris·HCl, pH 7.8; 50 mM NaCl; 10 mM $MgCl_2$; 1 mM dithiothreitol; bovine serum albumin at 100 μg/ml; 5 μM dATP; 5 μM dGTP; 0.5 μM biotin-11-dUTP (Enzo); 0.5 μM biotin-11-dCTP (Enzo); 1 unit of avian myeloblastosis virus reverse transcriptase; and the desired restriction enzyme or enzymes. The reaction mixture was incubated at 37°C for 2 hr. After precipitation with ethanol from 2.5 M ammonium acetate, the DNA was dissolved in 16 μl of 10 mM Tris·HCl, pH 8.0/1 mM EDTA. After incubation at 68°C for 10 min, secondary cleavage was performed by incubating with *Fok* I in the recommended buffer at 37°C for 3 hr. After the addition of 50 μl of 0.15 M NaCl/0.015 M sodium citrate/0.1% sodium dodecyl sulfate, the sample was passed over a spun Sephadex G-50 column (Select-B columns for biotinylated DNA from 5 Prime → 3 Prime). To attach any biotinylated DNA fragments to a solid support, the following were added to the column eluate: 75 μl of Triton wash (0.17% Triton X-100/100 mM NaCl/10 mM Tris·HCl, pH 7.5/1 mM EDTA) and 10 μl of a 0.4% suspension of streptavidin–CrO_2 particles. These particles are CrO_2 particles that have been coated with silica, treated with γ-aminopropyltriethoxysilane, and had streptavidin attached by glutaraldehyde crosslinking (7). The streptavidin–CrO_2 particles were washed three times in Triton wash just prior to use. The sample was incubated at 37°C for 30 min with occasional agitation. The sample tube was placed in a magnetic rack (Magic separator, Corning), causing the particles to become immobilized along the side of the tube. The liquid was removed by using a pipettor, the tube was removed from the rack, and the particles were resuspended in 100 μl of Triton wash. This washing procedure was repeated two more times. After the final wash, the particles were resuspended in 9 μl of 10 mM Tris·HCl, pH 8.0/1 mM EDTA. Fluorescence labeling of the *Fok* I ends was performed as described for direct fingerprinting. The particles were washed three times with 100 μl of Triton wash, using the magnetic rack as above. After the final wash, the particles were resuspended in 5 μl of 95% formamide/12.5 mM EDTA, incubated in a boiling water bath for 5 min, and placed in the magnetic rack. The liquid containing the eluted DNA strands was transferred to a fresh tube. Two microliters of the sample was loaded on a denaturing polyacrylamide gel, electrophoresed, and analyzed as described for direct fingerprinting.

RESULTS

Critical Reagents. This fingerprinting method takes advantage of the four fluorescent dideoxynucleotides (SF-ddNTPs) developed for automated DNA sequencing (6). Because each nucleotide is labeled with a slightly different succinylfluorescein dye, these molecules are nucleotide-specific reporters that permit sequence information to be determined directly from fluorescent emissions. The SF-ddNTPs are accepted by a number of DNA polymerases, making it easy to attach these nucleotide-specific reporters to the 3′ ends of DNA fragments.

The other critical reagent is a type of restriction enzyme that leaves a 5′ overhang where the sequence of the 5′ overhang is not unique and can consist of several different nucleotide combinations. Examples of these "5′-ambiguous end" restriction enzymes are shown in Fig. 1. The most important enzymes, exemplified by *Fok* I, are class IIS restriction enzymes that leave 5′ overhangs. Class IIS restriction enzymes do not cleave within the recognition sequence, but at a precisely defined distance away from it (11). Other enzymes that could be used are those like *Tth*111I that have a split recognition sequence and cleave in the middle to leave an ambiguous 5′ overhang or those like *Acc* I that have ambiguity within the recognition sequence.

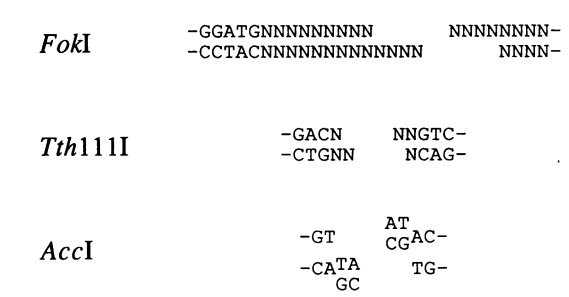

FIG. 1. Examples of 5′-ambiguous end restriction enzymes.

Direct Fingerprinting. A simple protocol for fingerprinting by sampled sequencing is outlined in Fig. 2. It is basically an extension of the method of Coulson *et al.* (1). In the primary cleavage step, each DNA sample is cleaved with a 5′-ambiguous end restriction enzyme. In the labeling step, a DNA polymerase is used to attach SF-ddNTPs complementary to the bases in the 5′ overhangs. If SF-ddNTPs are used alone, then only one fluorescent nucleotide will be added per end. The addition of unlabeled deoxynucleoside triphosphates (dNTPs) at the proper concentrations will allow the appropriate level of incorporation of fluorescent label at each of the positions of the overhang. Then, in the secondary cleavage step, the fluorescently labeled fragments are digested with a restriction enzyme or enzymes that are different from the enzyme used initially. This secondary cleavage serves several purposes. The first step labels both strands of each DNA fragment, which would prevent the determination of the sequence at either end. Secondary cleavage generates shorter fragments so that, in general, each labeled end is on a separate fragment of distinct size. Furthermore, the shorter fragments can be separated by procedures with single-base resolution, such as electrophoresis through denaturing polyacrylamide gels.

The DNA fragments are analyzed by electrophoresis through a denaturing polyacrylamide gel and the nucleotide-specific fluorescent dyes are discriminated using DuPont's Genesis 2000 system (6). With this system, the time taken for a labeled fragment to reach the detection zone is a measure of the mobility of that fragment through the gel. By comparing this with the mobility of known size standards, the size of each labeled fragment can be determined. Within the detection zone, the DNA fragments are irradiated by a laser beam

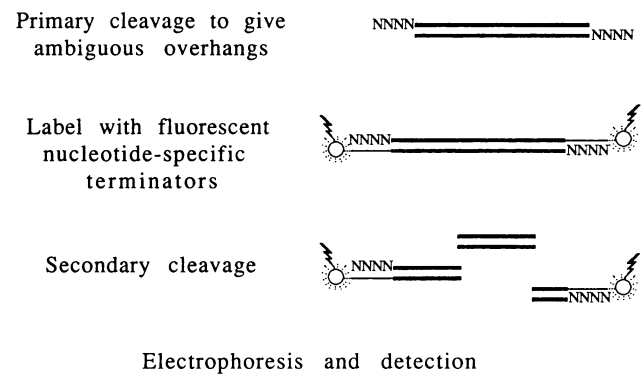

FIG. 2. Direct sampled sequence fingerprinting procedure. The Ns indicate the four-base 5′ overhangs generated by cleavage with *Fok* I.

and excitation/emission of the fluorescent dyes occurs as the fragments move through the zone. By using appropriate filters and a dual-detector system, each of the four fluorescence-labeled dideoxynucleotides can be identified on the basis of its distinctive emission spectrum. This information identifies the terminal sequence for each of the labeled fragments detected.

The feasibility of this simple labeling scheme was initially demonstrated by analyzing ϕX174 replicative form DNA. *Fok* I was used for primary cleavage and the ends were labeled by using reverse transcriptase to fill in the 5′ overhangs with a mixture of SF-ddNTPs and unlabeled dNTPs. After secondary cleavage with *Hin*fI, the labeled fragments were analyzed on the Genesis 2000. Representative results are shown in Fig. 3. Quartets of labeled fragments, with a spacing of one nucleotide within each quartet, are observed because *Fok* I leaves four-base 5′ overhangs. All of the labeled fragments expected on the basis of the known sequence of ϕX174 DNA were found. Where these labeled fragments did not overlap one another, the four-base terminal sequence could be determined. In each case, the terminal sequence was as predicted from the ϕX174 sequence.

Capture Fingerprinting. One limitation of the direct fingerprinting procedure outlined in Fig. 2 is that a 5′-ambiguous end restriction enzyme was used in the primary cleavage step. For a large DNA, there would be too many such sites and this would give a fingerprint in which sequence determination would be impossible. Optimally, a fingerprint should not have more than about 15 quartets, to avoid this problem. The versatility of sampled sequence fingerprinting has been greatly enhanced by the development of the procedure outlined in Fig. 4. In this fingerprinting procedure, any restriction enzyme or other method of specific cleavage is used in the primary cleavage step. Then, biotin is attached to each primary cleavage end either by using a DNA polymerase to fill in 5′ overhangs with biotinylated nucleotides or by using terminal transferase to add biotinylated nucleotides to blunt or 3′-overhang ends. Secondary cleavage is performed by using a 5′-ambiguous end restriction enzyme such as *Fok* I. Next, the DNA fragments are incubated with avidin or streptavidin immobilized on a solid support. DNA fragments with at least one biotinylated primary cleavage end bind to the immobilized avidin or streptavidin and all other fragments are washed away. The bound fragments are then

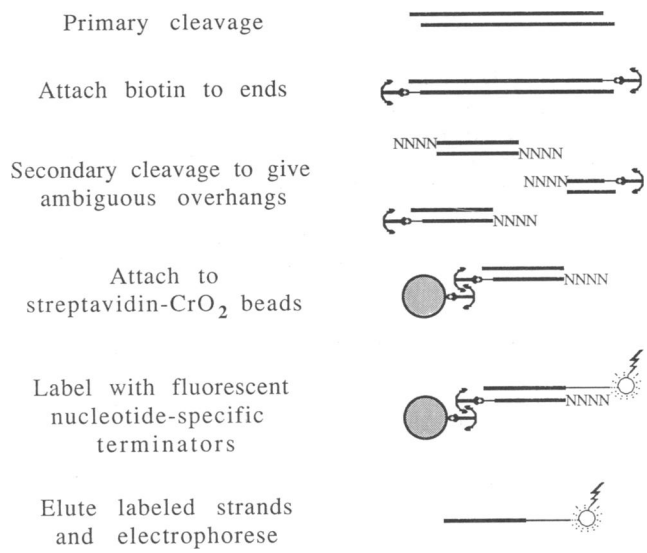

FIG. 4. Capture procedure for sampled sequence fingerprinting.

incubated with reverse transcriptase and SF-ddNTPs (with or without unlabeled dNTPs) to label the 5′-ambiguous ends generated by the secondary cleavage. Because the biotinylated nucleotides are added to the 3′ strand of the primary cleavage ends and the SF-ddNTPs are added to the 3′ strand of the secondary cleavage ends, the biotin and fluorescence label are on opposite strands for each labeled fragment. Thus, the strands labeled with fluorescence are eluted from the solid support by treatment with a denaturant such as formamide. The eluted labeled strands are then analyzed on the Genesis 2000 as before.

Fig. 5 shows representative results when this procedure was used to analyze phage λ DNA. Primary cleavage was accomplished by digestion with both *Bst*EII and *Bgl* II. Biotin was attached to the *Bst*EII and *Bgl* II ends by using biotinylated nucleotides as substrates for a polymerase fill-in reaction. After secondary cleavage with *Fok* I, fragments having a biotinylated *Bst*EII or *Bgl* II end were bound to streptavidin cross-linked to CrO_2 particles. Any *Fok* I ends of the bound fragments were labeled with fluorescence by using reverse transcriptase to fill in the 5′ overhangs with a mixture

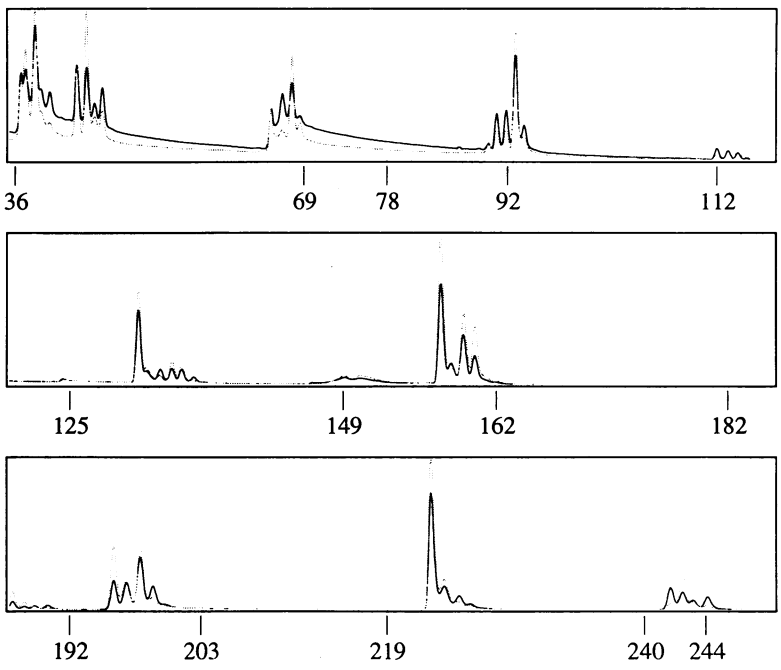

FIG. 3. Sampled sequence fingerprint of ϕX174 replicative form DNA. The ordinate represents fluorescence signal intensity as measured by the two photomultipliers (distinguished by solid and dotted lines) of the Genesis 2000. The abscissa represents electrophoretic migration time. The windows read from left to right and top to bottom, with each window showing 1 hr of electrophoresis time. Below each window is shown the position and size, in nucleotides, of marker DNA fragments electrophoresed in a parallel lane. The marker DNA was a pBR322 *Msp* I digest (New England Biolabs) that had been labeled by filling in the ends with dCTP and SF-505-ddGTP. On the basis of the published sequence of ϕX174 DNA (12), quartets of peaks of the following size and sequence are expected for the size range shown: 38–41, GGCT; 38–41, ACAA; 45–48, AGCC; 91–94, CCAC; 113–116, CTGG; 131–134, AATG; 159–162, GTGG; 159–162, ACCG; 199–202, CATT; 199–202, GATT; 225–228, AATC; 243–246, TGCT.

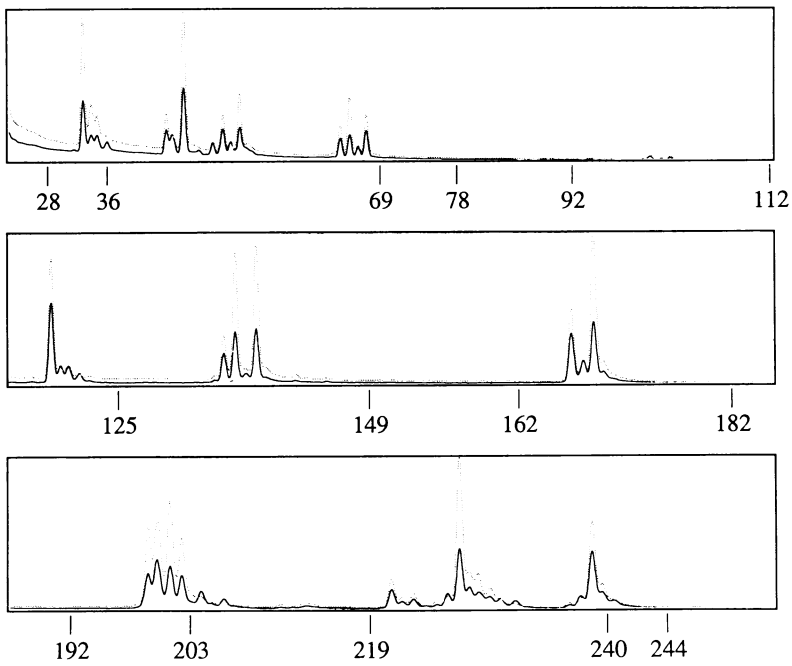

FIG. 5. Sampled sequence fingerprint of λ DNA. This fingerprint was prepared with 1.2 μg of λ DNA by the capture method. Primary cleavage was with a combination of *Bst*EII and *Bam*HI. Secondary cleavage was with *Fok* I. As in Fig. 3, fluorescence signal intensities are plotted versus electrophoresis time. Each window shows 1 hr of electrophoresis time. The position and size of pBR322 *Msp* I fragments electrophoresed in a parallel lane are shown below each window. On the basis of the published sequence of λ DNA (13), quartets of peaks of the following size and sequence are expected for the size range shown: 34–37, GGAT; 45–48, ACGC; 46–49, CGCT; 51–54, CCTT; 52–55, AAGA; 54–57, TTCT; 66–69, AGCA; 118–121, ATCT; 136–139, AGTC; 169–172, ATGA; 201–204, CAGG; 202–205, GGGA; 206–209, TCTC; 223–226, AAAT; 228–231, CGAT; 240–243, CAAC.

of SF-ddNTPs and unlabeled dNTPs. The fluorescence-labeled strands were eluted from the streptavidin-coated CrO_2 particles and analyzed by electrophoresis through a denaturing polyacrylamide gel using the Genesis 2000 detection system. In the size range of less than 300 nucleotides, all of the labeled fragments predicted from the known sequence of λ DNA were observed. Where fragments did not overlap, the correct terminal sequence was observed for each of the *Fok* I ends.

Comparison of Cosmids. We have examined the suitability of the capture method by analyzing cosmids from an ordered library of the bacterium *Rhodobacter capsulatus* constructed by J. Williams using the radioactive fingerprinting procedure of Coulson *et al.* (1). The comparison of three cosmids is shown in Fig. 6. Each cosmid is characterized by a distinctive set of patterns. The vector, in this case Lorist2, provides a set of patterns that are observed in every fingerprint. These vector patterns provide internal controls that can be used to normalize mobility results for fragments analyzed on different gels. Comparison of the other patterns shows that the three cosmids have a number of fragments of common size and terminal sequence, clearly indicating that these cosmids overlap. Cosmids A380 and A307 overlap extensively, whereas there is only a slight overlap between cosmid A797 and the other two. In comparison with the radioactive fingerprinting method, the most striking feature of sampled sequence fingerprinting is that finding a match is much more clear-cut and definitive.

DISCUSSION

The chief advantage of fingerprinting by sampled sequencing is that it increases the amount of information collected from each clone. As in previous methods, fragments up to at least 256 nucleotides are well resolved, so that size designation provides 8 bits of information. Designation of a four-base terminal sequence provides another 8 bits of information because there are 256 possible sequence combinations. Thus, each quartet pattern in a sampled sequence fingerprint provides 16 bits of information. Coincidence of two such frag-

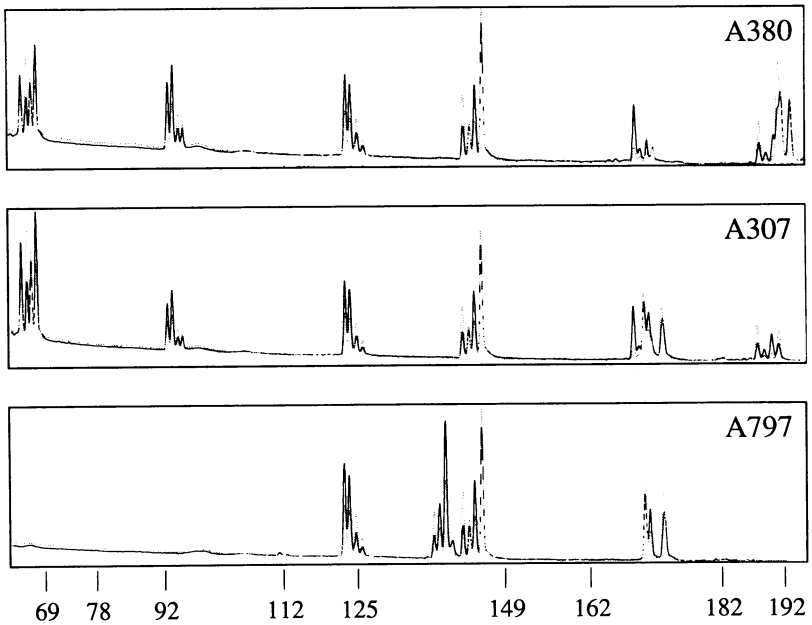

FIG. 6. Sampled sequence fingerprints comparing three cosmids with inserts of *Rhodobacter capsulatus* DNA. Fingerprints were prepared by the capture method, using *Bgl* II and *Hin*dIII as the primary cleavage enzymes and *Fok* I as the secondary cleavage enzyme. Fluorescence signal intensities are plotted versus electrophoresis time. Each window shows 2 hr of electrophoresis time. At the bottom are shown the position and size of pBR322 *Msp* I marker fragments electrophoresed in a parallel lane. The quartet at approximately 145 bases observed in all three fingerprints is derived from the vector, Lorist2, used to construct these cosmids. Cosmid DNA was prepared from 1.5 ml of overnight culture grown in "terrific broth" (8) by the alkaline lysis procedure (9, 10) with the following modifications. After the extraction with phenol/chloroform/isoamyl alcohol (50:49:1, vol/vol), nucleic acids were precipitated by addition of 460 μl of isopropyl alcohol. After RNase treatment, each sample was extracted with 60 μl of phenol/chloroform/isoamyl alcohol (50:49:1) and passed over a spun Sephadex G-50 column. The column eluate (approximately 70 μl) containing the cosmid DNA was stored at 4°C. Ten microliters of a cosmid DNA sample was used to prepare each fluorescent fingerprint.

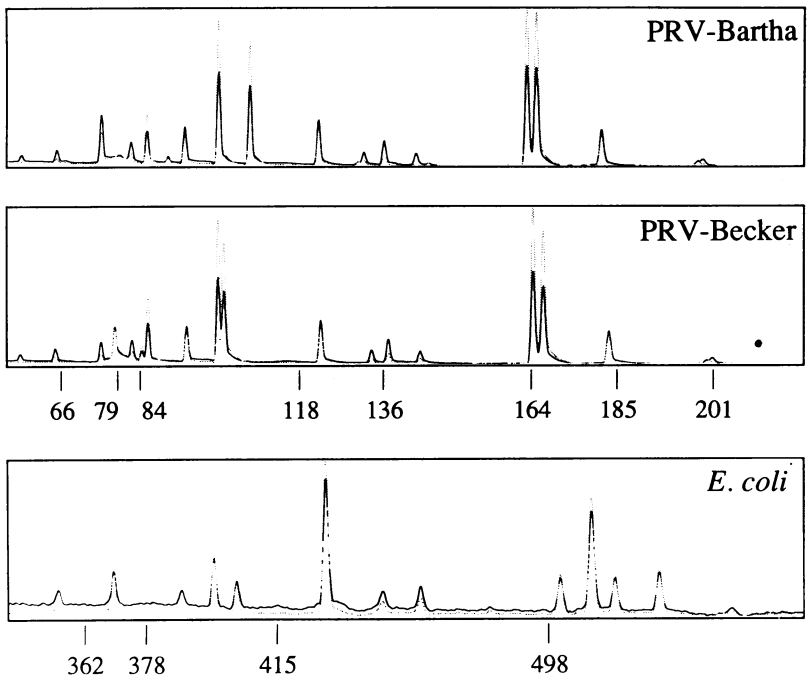

FIG. 7. Sampled sequence fingerprints of large DNAs. The DNAs analyzed are from two strains, Bartha and Becker (15), of pseudorabies virus (PRV), a herpesvirus that infects pigs, and from *E. coli*. The fingerprints were prepared with 0.5 μg of pseudorabies virus DNA or 6 μg of *E. coli* DNA by the capture method with the following changes. For pseudorabies virus DNA, primary cleavage was with *Sal* I; for *E. coli* DNA, primary cleavage was with *Xba* I. For all three samples, the labeling reactions contained SF-ddNTPs without any dNTPs present. Thus, only a single labeled nucleotide was added to each *Fok* I end, resulting in single peaks rather than the quartets observed earlier. The SF-ddNTP mix added to each of the labeling reactions contained 37.5 μM of each of the four SF-ddNTPs. These samples were electrophoresed on shorter gels than previously used, 20 cm as compared to 40 cm, and at lower power, 13 W. The pseudorabies virus samples were run on a different gel than the *E. coli* sample. Fluorescence signal intensities are plotted versus electrophoresis time. Each window shows 1 hr of electrophoresis time. The position and size, in nucleotides, of marker fragments are shown. On these gels, the size markers were λ *Bst*EII–*Fok* I and *Bam*HI–*Fok* I fragments.

ments would certainly establish overlap between fragments from a bacterial genome and be indicative for human genome clones.

We have noticed that each particular quartet has a distinctive and reproducible pattern of relative peak heights and spacing. This is likely to be an effect of the adjacent sequence on the efficiency of incorporation of the terminators and on the mobility of fragments in the gel. Because of these context influences, the exact pattern actually contributes more information than the tetranucleotide sequence itself.

Sampled sequence fingerprinting shares many advantages with the fluorescence-based fingerprinting method of Carrano *et al.* (14). In both methods, DNA fragments are detected at a fixed distance from the origin, leading to more uniform spacing between fragments than with conventional detection. Improved resolution is also realized because the fragments are detected by fluorescence rather than radioactivity. Both methods use automated data collection, which facilitates rapid and error-free data analysis and storage. However, the Carrano *et al.* method relies solely on size to characterize DNA fragments and thus suffers from limited information content. Carrano *et al.* can increase the information content per sample by running up to four separate digests in one lane. By including terminal sequence, our method of sampled sequence fingerprinting has a much greater information content per fingerprint with only a single fingerprint loaded per lane.

Another advantage of sampled sequence fingerprinting is that the capture method can be used to analyze very large DNAs. Gels have a limit to the amount of DNA that can be loaded. For example, Carrano *et al.* (14) report that loading more than 2 μg of DNA in a single lane of a denaturing polyacrylamide gel results in peak broadening and decreased resolution. In the capture method we determine the sequence and distance of the *Fok* I sites nearest the primary cleavage site. All the remaining DNA is eliminated by the washes after the streptavidin capture of biotinylated fragments. In principle, therefore, DNA of any length could be analyzed provided it has 10–20 primary cleavage sites. As examples of the analysis of large DNA, we show in Fig. 7 fingerprints of a herpesvirus of 150 kilobases and also a fingerprint of total *E. coli* DNA, which is 4.5 megabases in size. In these fingerprints only the first base of the quartet is labeled. These results are an indication of the versatility of fingerprinting by sampled sequencing. This method may therefore be most useful for characterizing large DNA clones and viral and bacterial genomes in the megabase range of size.

We are grateful to Terri McGuigan and Peter Korolkoff for providing technical assistance; the Diagnostic Systems group of the Medical Products Department, E. I. du Pont, for providing the streptavidin–CrO_2 particles; JoAnn Williams for providing the cosmid-containing strains; Lynn Enquist, Alan Robbins, and Mary Whealy for providing the pseudorabies virus DNA; and Nat Sternberg for providing the *E. coli* DNA. This is contribution no. 5153 from the Central Research and Development Department, E. I. du Pont.

1. Coulson, A., Sulston, J., Brenner, S. & Karn, J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7821–7825.
2. Olson, M. V., Dutchik, J. E., Graham, M. Y., Brodeur, G. M., Helms, C., Frank, M., MacCollin, M., Schein, R. & Frank, T. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7826–7830.
3. Daniels, D. L. & Blattner, F. R. (1987) *Nature (London)* **325**, 831–832.
4. Kohara, Y., Akiyama, K. & Isono, K. (1987) *Cell* **50**, 495–508.
5. Lander, E. S. & Waterman, M. S. (1988) *Genomics* **2**, 231–239.
6. Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., Cocuzza, A. J., Jensen, M. A. & Baumeister, K. (1987) *Science* **238**, 336–341.
7. Birkmeyer, R. C., Diaco, R., Hutson, D. K., Lau, H. P., Miller, W. K., Neelkantan, N. V., Pankratz, T. J., Tseng, S. Y., Vickery, D. K. & Yang, E. K. (1987) *Clin. Chem.* **33**, 1543–1547.
8. Tartof, K. D. & Hobbs, C. A. (1988) *Gene* **67**, 169–182.
9. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.
10. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 368–369.
11. Szybalski, W. (1985) *Gene* **40**, 169–173.
12. Sanger, F., Coulson, A. R., Friedmann, T., Air, G. M., Barrell, B. G., Brown, N. L., Fiddes, J. C., Hutchison, C. A., III, Slocombe, P. M. & Smith, M. (1978) *J. Mol. Biol.* **125**, 225–246.
13. Daniels, D. L., Schroeder, J. L., Blattner, F. R., Szybalski, W. & Sanger, F. (1983) in *Lambda II*, eds. Hendrix, R. W., Roberts, J. W., Stahl, F. W. & Weisberg, R. A. (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 519–676.
14. Carrano, A. V., Lamerdin, J., Ashworth, L. K., Watkins, B., Branscomb, E., Slezak, T., Raff, M., de Jong, P. J., Keith, D., McBride, L., Meister, S. & Kronick, M. (1989) *Genomics* **4**, 129–136.
15. Robbins, A. K., Ryan, J. P., Whealy, M. E. & Enquist, L. W. (1989) *J. Virol.* **63**, 250–258.