

Multiplex DNA Sequencing

GEORGE M. CHURCH AND STEPHEN KIEFFER-HIGGINS

The increasing demand for DNA sequences can be met by replacement of each DNA sample in a device with a mixture of N samples so that the normal throughput is increased by a factor of N . Such a method is described. In order to separate the sequence information at the end of the processing, the DNA molecules of interest are ligated to a set of oligonucleotide "tags" at the beginning. The tagged DNA molecules are pooled, amplified, and chemically fragmented in 96-well plates. The resulting reaction products are fractionated by size on sequencing gels and transferred to nylon membranes. These membranes are then probed as many times as there are types of tags in the original pools, producing, in each cycle of probing, autoradiographs similar to those from standard DNA sequencing methods. Thus, each reaction and gel yields a quantity of data equivalent to that obtained from conventional reactions and gels multiplied by the number of probes used. To date, even after 50 successive probings, the original signal strength and the image quality are retained, an indication that the upper limit for the number of reprobings may be considerably higher.

AT PRESENT, DNA SEQUENCING PROJECTS CONSIST OF THE selection of DNA clones (1-3), the treatment of these individually to base-specific reactions (4, 5), the separation of the reaction products by size on sequencing gels, and the computer-aided reading of separated bands as a sequence of A, C, G, and T bases (6-10). In that each DNA sample must be processed individually, a sequencing project entails numerous repetitions of each step. Multiplex DNA sequencing reduces the number of repetitive steps by the mixing together of different DNA samples. This "multiplexing" is analogous to signal multiplexing in electronics where several signals travel through a single channel simultaneously. The advantage is greatest when the mixing occurs as early as possible and separation occurs as late as possible. In our protocol, mixing occurs as soon as plasmid-bearing colonies are obtained from the initial recombinant DNA libraries (Fig. 1), and separation is done after the sequencing gel step. The separation process takes advantage of the ease of visualizing, by probing, one DNA sequence ladder among many latent ladders immobilized on nylon membranes as in the "genomic sequencing" method (11). In multiplex sequencing the probes are specific for tags contained within the vectors used to clone the DNA.

The genomic sequencing method was developed for studies of DNA methylation and DNA-protein interactions in living cells. It is inefficient for routine DNA sequencing for three reasons. (i) Each

sequencing ladder of 300 bases requires its own probe. (ii) Each ladder requires a properly located restriction site. (iii) The signal strength is limited by the low copy number of each target DNA sequence in natural genomes. Multiplex sequencing has eliminated these impediments by forming artificial genomes from mixtures of sonicated genomic fragments attached to a predetermined set of "tag" oligonucleotides.

In principle, multiplexing could combine clone selection by any nested deletion (3), complementary DNA (12), shotgun (1, 2), or DNA hybridization (1, 13) method with amplification by any host, vector, or polymerase chain reaction (14) system. Fragmentation could be done chemically (4) or enzymatically (5, 15). Finally, hybridization could use radioactively labeled or nonradioactive (16) probes. We describe one combination, consisting of shotgun selection, plasmid amplification, chemical sequencing, and ^{32}P labeling. The feasibility of other combinations remains to be determined.

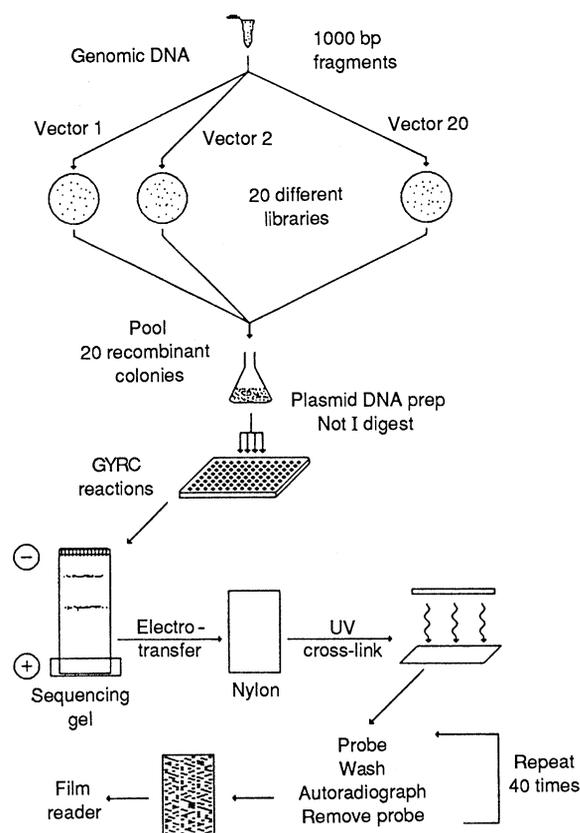


Fig. 1. The steps of multiplex sequencing. In order from the top, the diagram shows: total genomic DNA in a single tube, 20 libraries represented by colonies on 20 tetracycline-containing petri plates, a flask containing 20 pooled colonies in growth medium, a 96-well plate full of reactions, gel electrophoresis, a nylon membrane, the ultraviolet (UV) cross-linking step, and finally one of the 40 autoradiographs obtained by probing the membrane.

The authors are in the Department of Genetics, Harvard Medical School, and in the Howard Hughes Medical Institute, 25 Shattuck Street, Boston, MA 02115.

Fig. 2. Vectors for multiplex sequencing. A pair of tags flanks each Sma I site. For the plex00-20 series, *E. coli* rpoC terminator (30) containing oligonucleotides (upper DNA sequence) were cloned between Pvu II and Eco RI sites (31) destroying both recognition sequences. The resulting plasmid was cleaved at its new Eco RI and Pst I sites (in figure) for ligation to the mixed oligonucleotides shown in the lower DNA sequence. The nucleotides synthesized chemically are shown as uppercase letters and those filled in as lowercase. The symbols H25 and H26 represent all possible combinations of 25 and 26 A, C, and T nucleotides (32) that should have been present on these strands at these points after their chemical synthesis as a mixture. The symbols d25 and d26 represent the complementary sequences synthesized enzymatically. The numbers at the far right represent the sizes of the major oligonucleotide products at the states (left to right) of chemical synthesis, filling in, and cleavage. The 57-nt oligonucleotides were annealed to each other, and the single-stranded regions were filled in with the use of DNA polymerase I large fragment. The resulting mixtures of double-stranded 99-bp fragments were cut with Eco RI and Pst I and ligated into the plasmid above. The ligation products were transformed into *E. coli* strain DH5 (18). Randomly chosen clones were checked for the appropriate restriction sites and then sequenced in both directions by a simple version of multiplex sequencing method (one plasmid and two probes) with the use of the Hind III and Hinf I sites shown and probes proximal to those sites. Only one of the vector's eight Hinf I sites is shown. Regions forming 7- and 10-bp hairpin stems indicated above the upper DNA sequences by arrowhead pointing to the center of symmetry cause electrophoretic compression anomalies. These regions were again sequenced via bisulfite chemistry (22).

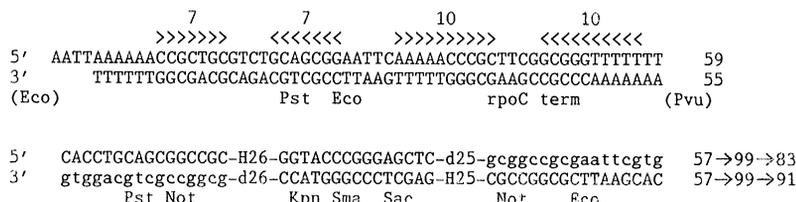
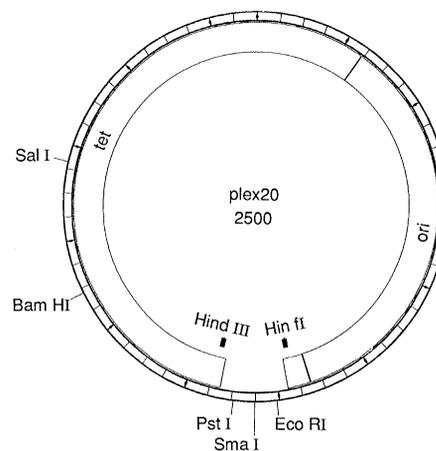


Fig. 3. The sequences of the probe hybridizing regions of the plex00-20 vectors. The first 20 nucleotides from each of the 40 sequences were synthesized and tested as probes. The 5' terminal (extreme leftmost) nucleotide of each is the 3' terminal cytosine of the Not I recognition sequence (5' GCGGCCGC 3'). The letter P indicates that the probe is closest to the Pst I site; the letter E, closest to the Eco RI site. The numbers represent both the vector and probe numbers. The dot at the end of the first line (01 P) indicates that this vector was found to be one nucleotide shorter than the others in that region (33). Number 00 is the standard plasmid used to provide an internal control of known sequence in every pool of unknown recombinants. An alternative probe to obtain internal standards from the vector sequences common to all of the vectors overlaps the Eco RI site and reveals the sequence extending from the 3' end of the Eco RI proximal Not I site counterclockwise through the *ori* sequence. The sequence of probe was GCGAATTCAAAAACCGCTT.

P 00	CCTTCCTAAACCACTCCA TTTAACC	E 00	CTTCCTTTTATCTTTATTA ACATTT
P 01	CCTTTCATTACAACATAATCA TAACAT.	E 01	CCCCCAATAAAATCATACTA CCTTAT
P 02	CATAAATTATTCTACTTATA AAAAATA	E 02	CTTTTACACAATAACTCTTA CTCAAA
P 03	CAACTCCTTCAACTCTACA CTTACTC	E 03	CTAACACAACACTTACTAC ATCTCA
P 04	CAATATTTTAAACCTCACACT TCCTAATA	E 04	CAACCCCATCCACTAACT TAACAA
P 05	CATAATATAACCTAAACCTT AAATCTT	E 05	CACATAACTCAAACTCTCAA TTCACC
P 06	CCACATCCAAAATAATCAAT CAACATA	E 06	CACCCATATCAATTTACAA TAATCT
P 07	CTACTAAATTTCTTTTATA TCCCCAA	E 07	CCATAAACTCCTCATCTTCC CCCTC
P 08	CCCTCCAATATAATATATA ATTACAT	E 08	CCTCTCCCATTTTTATTAT TACTTC
P 09	CATTAAACAATCATACCACTA CCAAATA	E 09	CTATTACAAATACACCACTA CAATCC
P 10	CCTAATCATCAATATATCTCA ATACAAC	E 10	CACCTAATACCTATATATA ATACCA
P 11	CATCACCACACATATTATCC TTATCCT	E 11	CGACCTTATACTTCTTACA ACTCAT
P 12	CTAATCTTAAACCACTTAA AATACCT	E 12	CCCTTCTACCAACCTATATC ATTTAT
P 13	CACTTACTCTACACCCCTT TTATCAC	E 13	CAAAAATAAATCCCAAAA ATCTCT
P 14	CTCTTGATAATATAATTAC TACTCCT	E 14	CTTCTTTTCTCAACCCCTT CACCCA
P 15	CAAACCAACATTTAACAACA TATATCA	E 15	CTCAATAATCTTATCTTAC TTTACA
P 16	CATAAACACCCATTTCATCCA ACTCTTA	E 16	CTTACAATCCCTATATATA CATTAT
P 17	CTCACCTTCTTAAACCCGAA ATTCTCT	E 17	CGAACCTCTTCAACCCGCAA TTTAAT
P 18	CAATCTACTTCCAACCACT ATTCAAA	E 18	CAACCAACAACCAACCTACT CCCCCTA
P 19	CAACCAACTCTACTTATGAC TCCCAAC	E 19	CCAACCTACATTAACCTCT ATCATC
P 20	CCCCACCAAAATCAACTAC AATACTC	E 20	CCATAACTTTAACCTTTAAC ATTAAT

The 20 new plasmid vectors used are simple and similar to one another (Fig. 2). In addition to standard drug resistance and replication origin elements, each has a cloning site flanked by two different oligonucleotide tags (40 total tags per 20 vectors). These tags are in turn flanked by rare restriction sites (Not I sites) so that the tagged DNA inserts can be easily excised after amplification; this excision produces the appropriate ends required for subsequent chemical sequencing. The vectors differ from each other only by their tag sequences, which were originally selected from a random collection of chemically synthesized oligonucleotides. The sequences of the tags were determined by multiplex sequencing (Figs. 2 and 3). These vectors were then used to clone genomic DNA as follows.

Genomic DNA (17) was sonicated (2) and sized on a native acrylamide gel (4), and the ends of the 900- to 1500-bp fractions were rendered ligatable by treatment with Bal 31 exonuclease, and then with T4 DNA polymerase (2) and all four deoxynucleotide triphosphates. These fragments were ligated separately into each of the 20 vectors, which had been treated with Sma I and alkaline phosphatase (2). These ligation mixtures were used to transform (18) *Escherichia coli* strain DH5. The resulting 20 libraries were amplified under tetracycline selection. The supercoiled DNA was fractionated on a 1 percent agarose gel to separate plasmids with single inserts away from those with zero, two, or more. The appropriate size fraction was excised, electroeluted, and used to

transform DH5 cells (18). Transformants were stored at -70°C as glycerol stocks.

Portions (10 to 150 μl) from each glycerol library stock were spread on tetracycline plates and incubated at 37°C for 24 hours, yielding about 100 colonies per plate. In order to obtain equal amounts of each plasmid in each pool, we used a 100- μl glass capillary tube to gently aspirate the central regions from 20 colonies of similar size, one from each library plate. All pools of mixed colonies were transferred from the capillaries to 40 ml of growth medium in 250-ml flasks and grown to saturation. Plasmid DNA was purified from each culture by alkaline lysis (19). After the first isopropanol precipitation, the pellets were resuspended in 150 μl of 50 mM tris-HCl, pH 8.3, then mixed with 100 μl of 8M ammonium acetate and transferred to 96-well plates (20). These plates were cooled to -70°C and held for 15 minutes; they were then centrifuged for 15 minutes at 3000g. The supernatants were transferred to new plates containing 150 μl of isopropanol per well. The plates were sealed (21), vortexed well, and centrifuged as before. The pellets were rinsed with ethanol, dried, and resuspended in 100 μl of 10 mM tris-HCl, pH 8.0, and 0.1 mM disodium EDTA. Portions of each DNA sample (20 μl ; about 20 μg) were cut with 20 units of Not I enzyme in 100 μl of buffer for several hours then precipitated twice with ethanol. As an internal standard, 1 μg of plex00 DNA was added to each sample.

Chemical cleavage reactions (4, 22) specific for guanines (G), purines (R), cytosines (C), and pyrimidines (Y) were modified as described below to allow the use of 96-well plates. Each DNA sample cut by Not I was resuspended in 35 μ l of water, and 5 μ l of each sample was delivered to each of two wells of one set of polypropylene 96-well plates (20) for G and R reactions and into two wells of a second set of plates for the C and Y reactions. To the R wells were added 10 μ l of R reagent. This R reagent consisted of 15 mM acetic acid in water. These plates were taped and incubated at 65°C for 15 to 50 minutes, and allowed to cool; then 10 μ l of G reagent was added to the G wells. This G reagent consisted of 1 mM dimethyl sulfate in ice-cold water; the dimethyl sulfate was diluted in cold water just before use since it immediately starts to decompose into methanol and sulfuric acid. The dimethyl sulfate and hydrazine should be handled in the hood until neutralized (4). These plates were further incubated at room temperature for 50 minutes and then 50 μ l of 1.3M piperidine was added to each well.

To the appropriate C and Y wells of the C and Y plates were added, with mixing, 15 μ l of C and Y reagents. The C reagent consisted of a mixture (10:5) of hydrazine and 3M sodium acetate in water. The Y reagent consisted of hydrazine and water (9 : 6). These C and Y plates were incubated for 15 minutes at room temperature (23°C). The reactions were stopped by the addition (with mixing)

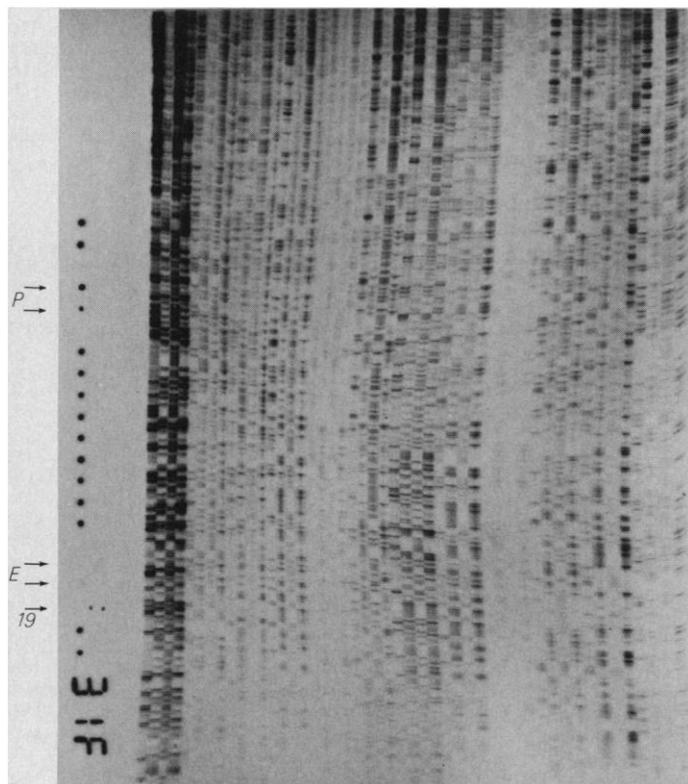


Fig. 4. An autoradiograph showing the standard format for reactions and DNA markings. Each membrane had 48 lanes, 12 sets of GYRC reactions (left to right), which is equivalent to one row (12 wells) from a 96-well clone pool storage plate (in this example, plate 31 row F). The identifying name (31F) was drawn on the left edge of the dry membrane prior to transfer, with a mixture of all of the vectors denatured in alkali. In this way, each probing yielded the name 31F regardless of the probe used. The membranes were also marked with three kinds of DNA dots. The first kind, placed in the regions indicated by the P (or E) arrows, were made with a pool of the P (or E) halves of all vectors. The second series of 20 replicate dots represents DNA from each of the 20 vectors. In the film pictured here, the pair of dots near the P and the pair of dots near the 19 indicate unambiguously that the probe used was P19. All remaining dots on the film were made with the same pool of vector DNA used to write the name of the membrane. They, too, light up with every probe and are used for alignment of the film in data analyses.

of 200 μ l of 300 mM sodium acetate and 1M acetic acid, pH 4.5, and then 150 μ l of isopropanol. The DNA samples were sedimented by a 15-minute centrifugation, and the pellet was resuspended in 200 μ l of 300 mM sodium acetate, pH 7. The samples were precipitated two more times with 150 μ l of isopropanol, rinsed with ethanol, centrifuged, dried, and resuspended in 65 μ l of 1M piperidine.

Next, all of the plates were taped, covered with a previously warmed glass sheet, incubated at 90°C in an oven for 40 minutes, and then cooled. The tape was removed, and the plates were then loosely covered with plastic sheets and incubated at -70°C for 20 minutes to freeze the solutions. The samples were lyophilized at reduced pressure (<0.1 mmHg) and resuspended in 10 μ l of water. The freeze-drying was repeated, and the samples were resuspended in 5 μ l of a solution of 50 percent formamide in water and heated to 90°C in a vacuum oven for 10 minutes. Finally, 1 μ l of each sample was loaded onto ionic gradient sequencing gels (23). After separation of the DNA reaction products by size on the gels, the DNA patterns were electrotransferred from the gels to nylon membranes (24) and cross-linked by ultraviolet light (11). To correlate and keep track of the films, samples, membranes, and probes throughout the subsequent 40 probings the membranes were marked, while dry before transfer, with DNA markings (Fig. 4).

Probes were made in the following manner. Oligonucleotides, 20 nt in length and complementary to the tag sequences in the vectors (Fig. 3), were chemically synthesized, gel-purified, eluted with water, and adjusted to 4 pmol/ μ l. Portions (1 μ l) were labeled by the addition of about 15 dCMP residues from [α -³²P]dCTP (5000 Ci/mmol; Du Pont) with calf thymus terminal deoxynucleotidyl transferase (25) at 0.3 U/ μ l in a mixture of 120 mM cacodylic acid (adjusted to pH 7.6 with potassium hydroxide), 1 mM β -mercaptoethanol, 2 mM CoCl₂, and bovine serum albumin (BSA) at 0.2 mg/ml and then incubating at 37°C for 30 minutes. For some reactions, 1 μ l or less was analyzed by electrophoresis on a small, 20 percent acrylamide sequencing gel to check that more than 60 percent of the ³²P label had been incorporated and to estimate the number of residues added. Labeling with T4 polynucleotide kinase (4) was used in our early experiments, but we had difficulties obtaining nearly complete incorporation of both the oligonucleotides and labeled phosphate into labeled oligonucleotide products. The 25- μ l terminal transferase reactions were diluted into 20 ml of hybridization buffer.

Two different probing methods were used. In the first method, up to 80 membranes were placed in a container (24) with 10 to 30 μ l of hybridization solution per square centimeter of membrane [7 percent SDS, 10 percent polyethylene glycol (PEG), 0.25M NaCl, and 0.13M phosphate buffer (26)]. These were incubated for 4 hours at 42°C and then washed at room temperature in plastic tanks with four 1000-ml portions of 0.07M phosphate buffer containing 1 percent SDS. Each of the moist membranes was placed between Saran wrap and 0.38-mm-thick polyester. After exposure to x-ray film for 1 to 10 hours at room temperature, the probes were removed with 0.1 percent SDS and 2 mM EDTA (adjusted to pH 8.3 with tris base) at 50°C for 1 to 10 minutes. More than 99 percent of the radioactivity remaining from the previous hybridization was removed by this procedure. With this and the alternative procedure below, resolution was essentially the same as in standard sequencing methodology (23). Signal strength, nonspecific binding, band resolution, and membrane dimensions do not change perceptibly through 44 probings (Fig. 5). This result has, so far, been extended to 50 probings. Probes E02, E08, and E15 (Fig. 3) gave generally weaker signals than the rest.

An alternative probing procedure in which pouches were recycled through the hybridization, washings, and autoradiography steps

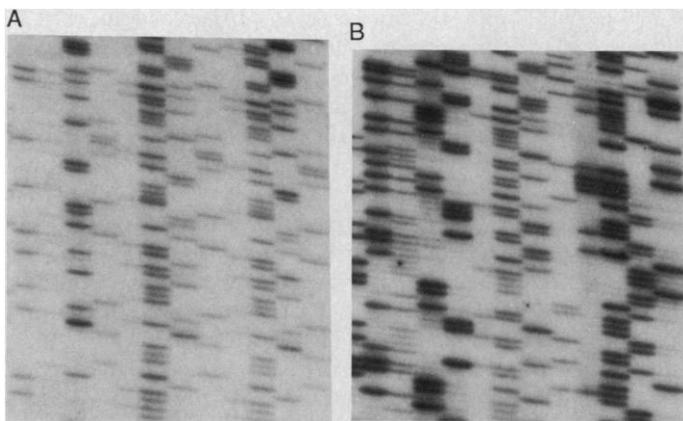


Fig. 5. Probing number 4 and number 44 of membrane 01A. The probes were E00 (A) and E09 (B), respectively. The order of chemical sequencing reactions from left to right is GRTCGRTCGRTC. The three sets of four reactions for the E00 standard display the same sequence, as expected. These read (3' to 5') from 9 to 16 cm from the bottom of the gel: GAGTACAACTGTCGAATAGTAGCTATTCGAAAATTACGCCATCAAATAGTGTCAATTTAAC. This covers bases 110 to 161 from the 3' end of the upper strand of the Not I site (rightmost Not I site on the lower DNA sequence of Fig. 2). This reads clockwise into the *tet* gene (30).

without disassembly appears to be more amenable to automation and eliminates membrane handling through the many cycles of probing. The sealed pouches, containing one membrane each, were stretched taut on horizontal acrylic plastic sheets with tape, and the hybridization and washing solutions were introduced through tubing (27). After the washings, the images were recorded by sliding x-ray film over the pouch surface and pressing the film, pouch, and membrane tightly together (28).

Automatic DNA sequencing film readers (7, 29) can benefit from two additional multiplexing features. Known DNA sequences (plex00, see above) included in each lane act as ideal internal standards, since the band and lane shapes and reaction chemistry visible from one probing are congruent with those in all subsequent probings. This aids the data reduction steps and the quantitative recognition of trouble spots. The DNA markings mentioned earlier identify the membrane number and probe number on each film by hybridization (Fig. 4). Their presence on the films helps in the computer-aided alignment of internal standard lanes with unread sequence lanes, the documentation and retrieval of the films, and the correlation of the film data to the stored DNA samples.

REFERENCES AND NOTES

1. F. Sanger, A. R. Coulson, G. F. Hong, D. F. Hill, G. B. Petersen, *J. Mol. Biol.* **162**, 729 (1982).
2. P. L. Deininger, *Anal. Biochem.* **129**, 216 (1983).
3. S. Henikoff, *Gene* **28**, 351 (1984).
4. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
5. F. Sanger, S. Nicklen, A. R. Coulson, *ibid.*, p. 5463.
6. R. Staden, *Nucleic Acids Res.* **12**, 499 (1984).
7. J. Elder, D. K. Green, E. M. Southern, *ibid.* **14**, 417 (1985). Sequence autoradiograph readers developed by Bio-Rad, Hitachi, and Seiko are mentioned in *Nature (London)* **325**, 771 (1987); *ibid.* **328**, 460 (1987).
8. C. Connell *et al.*, *BioTechniques* **5**, 342 (1987).
9. J. M. Prober *et al.*, *Science* **238**, 336 (1987).
10. R. Staden, *Nucleic Acids Res.* **10**, 4731 (1982).
11. G. M. Church and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1991 (1984).
12. T. V. Huynh, R. A. Young, R. W. Davis, in *DNA Cloning Techniques: A Practical*

Approach, D. Glover, Ed. (IRL Press, Oxford, 1985), p. 49; H. Okayama and P. Berg, *Mol. Cell. Biol.* **2**, 161 (1982).

13. S. M. Honigberg, B. J. Rao, C. M. Radding, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9586 (1986); B. Rigas, A. A. Welcher, D. C. Ward, S. M. Weissman, *ibid.*, p. 9591.
14. R. K. Saiki *et al.*, *Science* **230**, 1350 (1985).
15. Y. Kohara, K. Akiyama, K. Isono, *Cell* **50**, 495 (1987).
16. J. J. Leary, D. J. Brigati, D. C. Ward, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4045 (1983).
17. Single bacterial colonies of *Escherichia coli* K-12 EMG2 (*E. coli* Genetic Stock Center Yale University CGSC 4401) and *Salmonella typhimurium* LT2 (American Type Culture Collection ATCC 23564) were grown to saturation. Part of each culture was used to prepare DNA and the rest was stored at -70°C in 15 percent glycerol.
18. D. Hanahan, *J. Mol. Biol.* **166**, 557 (1983).
19. D. Ish-Horowicz and J. F. Burke, *Nucleic Acids Res.* **9**, 2989 (1981); T. J. Gibson and J. E. Sulston, *Gene Anal. Technol.* **4**, 41 (1987).
20. Polypropylene 96-well plates with V-bottom wells (Micronic Systems, Netherlands) were used because they do not warp at 90°C , the temperature required for the piperidine steps. The wells (cut to a height of 2.5 cm) hold 400 μl . For loading from such plates, we have used 12-channel pipettors with 25-mm-long needles (tapered, 26 to 31 gauge) spaced 9 mm apart (designed in collaboration with Hud Schultz and available as item 0155200, Hamilton, Reno, NV). The reactions were arranged within the 96-well plates so that after loading the samples would appear in the order GYRC. The centrifugations were done at top speed in an Omnifuge RT (American Scientific Products).
21. Adhesive tape (3M Scotch 3750-G) was placed directly on top of the 96-well plate sealing off each well.
22. For a few early sets (Fig. 5) we used the thymine-specific KMnO_4 reaction; C. M. Rubin and C. W. Schmid, *Nucleic Acids Res.* **8**, 4613 (1980). In some sets urea-bisulfite [N. Okada, K. Sakamoto, Y. Itoh, Y. Oshima, *J. Biochem.* **91**, 1281 (1982)] or methoxylamine-bisulfite [N. S. Ambartsumyan and A. M. Mazo, *FEBS Lett.* **114**, 265 (1980)] reactions were used to eliminate compressions.
23. M. D. Biggin, T. J. Gibson, G. F. Hong, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3963 (1983).
24. The nylon membranes (Gene Screen, Du Pont) were cut to 26 by 36 cm. The membranes expanded uniformly (up to 4 percent) on hydration and then varied by less than 0.2 percent thereafter. The electrotransfer devices measuring 11 by 31 by 42 cm (inner dimensions) were from Polytech Products, Somerville, MA. Hybridization containers were either acrylic tanks of the same dimensions or plastic pouches measuring 48 by 40 cm, composed of 60- μm -thick polyethylene-polyester laminated films obtained from Spec Fab, Riverton, NJ.
25. R. Roychoudhury and R. Wu, *Nucleic Acids Res.* **3**, 101 (1976).
26. R. M. Amasino, *Anal. Biochem.* **152**, 304 (1986).
27. S. W. Compton *et al.*, *BioTechniques* **3**, 308 (1985).
28. A pressure-regulated expanding air pouch was placed above the pouch housing the membrane and both pouches were confined to a space (56 by 48 by 4 cm) delimited by acrylic plastic. Exposure defects caused by static electricity were eliminated by the temporary placement of grounded, high-density, 6-mm-thick conductive foam (Foamfab, Mansfield, MA) over the x-ray film while sliding the film into place.
29. G. M. Church, unpublished data.
30. C. Squires, A. Krainer, G. Barry, W.-F. Shen, C. L. Squires, *Nucleic Acids Res.* **9**, 6827 (1981); T. J. Gibson, A. R. Coulson, J. E. Sulston, P. F. R. Little, *Gene* **53**, 275 (1987).
31. The first numbers given below after a site refer to the coordinates in the GenBank (TM) database and the second numbers to the corresponding points in the plex20 vector. The precursor plasmid backbone is composed of the pBR322 origin fragment Hae II (2353; 1126) to Dra I (3234; 244), the pBR322 *tet* gene fragment Ava I (1425; 243) filled in to Eco RI (4363; 1317), and the *lac i* gene fragment Pvu II (1122; 1178) to Hae II (1069; 1127). Preparation of Figs. 2 and 3 was aided by the University of Wisconsin Genetics Computer Group software; see J. Devereux, P. Haerberli, O. Smithies, *Nucleic Acids Res.* **12**, 387 (1984). The vector sequences have been deposited in GenBank (TM).
32. A. Cornish-Bowden, *ibid.* **13**, 3021 (1985).
33. The Not I site normally adjacent to the Pst I site is absent in plex00 (the fifth base was deleted) so that a sequencing ladder extending from the Not I site adjacent to the Eco RI site (Fig. 2), upon hybridization with probe E00, gives a 2499-nt-long set of standard sequence markers (3' end-labeled as are the unknown sequences). Vector plex06 lacks a Not I site on its Pst I end (an A instead of a G precedes the first C in the line for P06). Vector plex20 lacks its Kpn I site due to a T replacing the leftmost C (Fig. 2). Vectors plex11, -12, and -14 appear to have reduced Kpn I cutting accompanying a 5mC replacement of the same leftmost C.
34. We thank D. Agard, W. Gilbert, G. Martin, H. Martinez, C. Ordahl, J. Sedat, D. Wiley, T. Wu, and members of their labs for sharing facilities and thoughtful suggestions; B. Bachmann, T. Gibson, J. Roth, and R. Tizard for stimulating discussions and for advice on strains, vectors, and chemistry; L. Christiansen and G. Gryan for oligonucleotides; E. Burgess, J. Horn, and H. Schultz for help on apparatus construction. Supported by a Monsanto-LSRF fellowship, DOE grant DE-FG02-87ER60565, and the Howard Hughes Medical Institute.

5 October 1987; accepted 16 February 1988