JOURNAL OF BACTERIOLOGY, Mar. 1990, p. 1167–1172
0021-9193/90/031167-06$02.00/0

Vol. 172, No. 3

MINIREVIEW

New Approaches for Physical Mapping of Small Genomes

CASSANDRA L. SMITH[1]†* AND GUY CONDEMINE[2]‡

Departments of Microbiology,[1] Psychiatry,[1] and Genetics and Development,[2] College of Physicians and Surgeons, Columbia University, New York, New York 10032

INTRODUCTION

Although linear chromosomal DNAs of fungi and protozoa (up to about 10 megabases [Mb] in size) (10, 23, 35) are fractionated by pulsed-field gel electrophoresis (PFG; 28), circular bacterial chromosomal DNAs (1 to 20 Mb) are not. Thus, methods were developed that use agarose as a protective matrix (32) to purify intact bacterial chromosomes that were enzymatically convertible to linear molecules. The first application was the construction of a complete low-resolution *Not*I restriction map of the *Escherichia coli* chromosome (31). Recently some *Sfi*I and some *Avr*II sites were added to this map (8a).

The genetic map and the wealth of knowledge available for *E. coli* aided the construction of a physical map, but the methods developed are general and allow physical-map construction for virtually any organism. These methods yielded the first complete restriction map of a eucaryotic genome (*Schizosaccharomyces pombe*, 15 Mb in size [10]) and are being used to construct a map of the smallest human chromosome (50 Mb in size) (C. L. Smith et al., unpublished data). Here we describe the application of this new technology to mapping small genomes and how such maps can be used to locate genes on the chromosomes. Some approaches are limited to organisms, like *E. coli*, for which much sequence and genetic-mapping data are available. However, large-DNA techniques should soon provide physical maps and their applications for a wide number of microorganisms. This article is intended to explain how to start using the new technologies. Other articles contain more complete descriptions of our methods (29, 32, 34) and PFG theory (5) and a detailed analysis of PFG parameters (4, 19–21).

PFG AND LARGE-DNA TECHNOLOGY

Principles and practice of PFG. DNA molecules of <20 kilobases (kb) are fractionated by conventional electrophoresis in matrixes such as agarose because they are sieved as they move. DNA molecules of >20 kb all move through agarose at the same speed because they are not sieved. Fractionation of DNA by PFG occurs because the speed at which a DNA molecule can change direction is size dependent. Larger molecules change direction more slowly than smaller ones. In PFG, DNA molecules are forced to change direction by exposure to alternating electrical fields. The time each field is on is called the pulse time. It is the major, and usually the only, variable adjusted in typical experiments. In PFG experiments a pulse time is chosen so that a particular DNA size class spends most of each pulse changing direction (i.e., involved in a size-dependent process) rather than moving. Hence, separation is based on retarding the net movement of DNA molecules through gels, which explains the requirement for long run times.

Since the first description of PFG (28), many variations have been described which, unfortunately, have been given different names. They all work on the same principle. The major differences are how the electrical fields are alternated, the angle of the DNA direction changes, and the shape of the electrical fields. These differences affect the straightness and resolution of the lanes of fractionated DNA. Straight lanes may be obtained with any PFG apparatus (5) by suitable adjustment of electrode positions and running parameters, but methods that yield straight lanes cause a loss in resolution. Most physical-mapping experiments are helped more by maximum resolution than by straight lanes (8a, 16, 31). All of our experiments use a Pharmacia-LKB Pulsaphor apparatus that maximizes resolution and spreads DNA samples over 15 cm of gel.

The usual PFG running conditions are comparable to those used for ordinary electrophoresis (32). The pulse time, ranging from 1 to 100 s, is usually the only variable for DNA molecules of less than 1 Mb. Typical run times are 40 h. For DNAs larger than 1 Mb, lower field strengths, e.g., 3 V/cm, are used to minimize shear damage, pulse times range up to 1.5 h, and run times range up to 6 days. Note that changes in any PFG parameters may affect mobility and resolution markedly (20, 21). We recommend starting PFG work with *Saccharomyces cerevisiae* chromosomal DNAs. These samples are extremely simple and inexpensive to make (32), the chromosome sizes are known (20), and the PFG results are easy to interpret.

Preparation of chromosomal DNA. Intact chromosomal DNAs are prepared in agarose blocks, called inserts (usually 0.1 ml), to protect them from shear damage (29, 32, 34). These samples are extremely easy to prepare and can be used for conventional recombinant-DNA procedures if precautions are taken to prevent loss of small fragments. For new organisms the amount of cells per insert should be determined empirically. For *E. coli*, 10^8 cells yields about 0.7 μg of DNA; usually one-sixth of an insert, at a DNA concentration of 0.5 μg/0.1 ml, is loaded per gel lane. New bacterial samples should be made at 0.1 to 2.5 μg per insert, and the amount of DNA loaded per gel lane should be adjusted by taking different slices of the insert.

Our original chromosome purification procedure used cells incubated for 1 h in medium containing chloramphenicol before harvesting, which synchronizes DNA replication forks at the origin, since ongoing rounds of DNA replication are completed but no new rounds initiate (22). This alignment is not necessary, but it is preferred because it ensures

* Corresponding author.
† Present address: Human Genome Center MS1-213, Lawrence Berkeley Laboratory, and Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720.
‡ Present address: Laboratoire de Microbiologie, INSA, 69621 Villeurbanne, France.

equal yields of DNA from all regions of the genome. Otherwise, DNA near the replication terminus may be underrepresented (33).

DNA purification involves distributing a suspension of live cells in 0.5% low-gelling-temperature molten agarose into a mold, which upon cooling forms inserts of 2 by 5 by 10 mm. The inserts are pushed out of the mold and incubated overnight in EC (32) lysis solution. The bacterial inserts, or inserts made from any organism, are then incubated for 2 days in ESP (32) solution, which removes all protein and other materials bound to DNA. For restriction enzyme digests, inserts are (i) rinsed in buffer, (ii) incubated with restriction enzymes for 2 to 4 h, and (iii) treated again with ESP to remove restriction enzymes from the DNA before an appropriate slice of the insert is loaded into a PFG well.

Large DNA molecules are exquisitely sensitive to contaminating nucleases and metals. These samples must always be nuclease free during handling. All solutions should be autoclaved or filter sterilized. No metal of any kind should be used to handle any DNA sample. The insert samples can be manipulated with an alcohol flame-sterilized glass rod and sliced with a glass cover slip.

DNA-agarose preparations can have several problems. Many agarose batches contain contaminants that interfere with in situ enzymatic manipulations of DNA. Seakem (FMC) now sells insert grade agarose (InCert) and GTG agarose (a low endoosmosis agarose), which test free of enzymatic inhibitors. In general, a good batch of agarose will allow use of a wide variety of restriction enzymes and DNA-modifying enzymes. DNA can diffuse out of agarose. This is especially bothersome with pieces under 20 kb. Storing the samples in a minimal volume of liquid largely circumvents this problem. Small fragments can be recovered from the supernatant and melted insert sample by treatment with agarase.

The *E. coli* procedure has been also used for mycobacteria, salmonellae, legionellae, mycoplasmas, bacilli, archaebacteria, and many other microorganisms. Sometimes it is necessary to modify the procedure slightly. For instance, the salt concentration in EC lysis is lowered for use with halophilic organisms (27). In our experience, the reported shorter protocols have hindered PFG experiments by yielding lower-quality DNA preparations. We recommend practicing by purifying *E. coli* K-12 DNA and digesting it with *Not*I. Ethidium bromide staining of a PFG separation will allow assessment of the DNA preparation, digestion, and fractionation procedures (Fig. 1A).

GETTING STARTED

Sample concentration and restriction enzymes. In principle, the construction of genomic and plasmid restriction maps is the same, except that the former involves DNA molecules 100 times larger. The major variables in applying PFG to new organisms are determining the appropriate amount of DNA to load onto a gel and identifying restriction enzymes that will produce a small number of genomic restriction fragments. The best PFG results are obtained with the smallest amounts of DNA. The amount to use must be determined empirically because for a new organism, one does not know a priori the genome size, the number of genomes per cell, or the density of restriction sites. Thus, the first experiments with a new organism should involve both a DNA titration and a survey of various restriction enzymes.

The only two available restriction enzymes with 8-base-pair recognition sequences, *Not*I and *Sfi*I (Table 1) should cut a random sequence of 50% GC content every 64 kb. Since *Sfi*I does not cleave DNA within a specific sequence

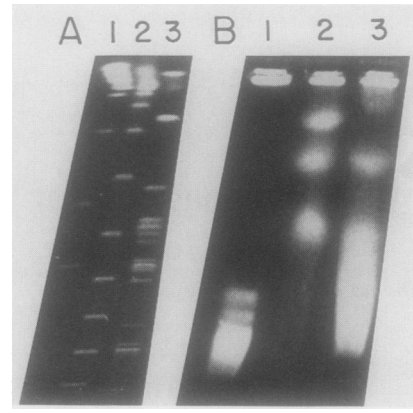

FIG. 1. Simple PFG analysis of small genomes. (A) *E. coli* chromosomal DNA cut with the restriction enzyme *Not*I (lane 3) and fractionated by PFG along with *S. cerevisiae* chromosomal DNAs (lane 1) and annealed λ phage DNA (lane 2). (B) Randomly broken full-sized linear *E. coli* chromosomal DNA (lane 3) was fractionated along with chromosomal DNAs from *S. cerevisiae* (lane 1) and *S. pombe* (lane 2). The *S. pombe* chromosomal DNAs are 3.5, 4.7, and 5.7 Mb in size (10). Pulse times for PFG fractionation were 25 s (10 V/cm) (A) or 3,600 s (3 V/cm) (B) for 48 h or 6 days, respectively (32).

(Table 1), *Sfi*I ends cannot be easily cloned, and the interpretation of *Sfi*I digests is complicated by the very different cleavage kinetics observed at different sites. *Not*I and *Sfi*I are not useful for GC-rich genomes; *Dra*I, *Ase*I (Smith et al., unpublished observations), and *Ssp*I (Table 1) may be used instead for such genomes.

Large fragments may sometimes be generated by other enzymes with 6-base recognition sequences. For some genomes it is effective to use *Eag*I or *Bgl*I, which recognize the inner six base pairs of the *Not*I and *Sfi*I restriction sites, respectively. Restriction enzyme sites containing the stop codon, TAG, appear very infrequently on the *E. coli* chromosome (8a). Of all enzymes tested, *Avr*II cleaves the *E. coli* genome into the fewest fragments. Although restriction enzymes recognizing other general transcription or translation regulatory sequences have not proven useful with *E. coli*, they may be useful with other organisms. Screening a wide variety of common enzymes against *Halobacterium* species revealed the usefulness of *Bam*HI (27). The obvious candidates listed in Table 1 or inferred from GC content

TABLE 1. Restriction enzymes useful for PFG experiments[a]

Enzyme	Recognition sequence[b]
*Not*I	GC/GGCCGC
*Sfi*I	GGCCNNN/NGGCC
*Dra*I	TTT/AAA
*Ase*I	ATT/AAT
*Ssp*I[c]	AAT/ATT
*Avr*II[c]	C/CTAGG
*Xba*I	T/CTAGA
*Spe*I	A/CTAGT
*Nhe*I	G/CTAGC
*Rsr*II	CG/G(A)CCG T

[a] Other restriction enzymes which may be useful are those containing the stop codon TGA (*Bsp*HI) or TAA (*Afl*II and *Hpa*I), the start codon ATG (*Nsi*I and *Sph*I), or a TATA sequence (*Ssp*I and *Eco*RV).

[b] Cutting sites are indicated by a slash. Biologically significant sequences are underlined. N, Any nucleotide.

[c] Many lots of *Ssp*I and *Avr*II are too contaminated with nonspecific nucleases to be useful for genomic mapping.

estimates should be tested first, but if necessary, a large variety of enzymes can be screened.

Determination of approximate GC content and genome size. The first PFG experiments will also characterize the chromosomal GC content. The average fragment size produced by cutting with a particular restriction enzyme is only an approximate measure of the GC content or dinucleotide frequency (8a, 28a, 31). Hence, the average size of fragments produced by a variety of enzymes allows only an estimate of the GC content.

Genomic size is determined by totalling restriction fragment sizes with several enzymes (8a, 16, 31). The number of fragments in unresolved multiplets may be resolved by varying pulse times or by densitometric quantitation. The lowest possible amounts of DNA should be loaded for correct sizing. The size discrepancy reported by various workers for *Mycoplasma* genomes (3, 24, 40) is probably due to overloading of the bacterial or size standard DNA samples and to the use of a reverse-field apparatus (field inversion gel electrophoresis [FIGE], see below).

Small fragments are difficult to detect because (i) they readily diffuse out of agarose and (ii) ethidium bromide staining is proportional to size. They can be detected by autoradiography of end-labeled genomic restriction fragments fractionated by short-pulse-time PFG or by conventional electrophoresis. DNA end labeling, in agarose, is effective with polynucleotide kinase or the Klenow fragment of DNA polymerase I (32).

DNA fragments smaller than 1 Mb can be sized by comparison with annealed bacteriophage λ size standards (50-kb monomer; 20). Finer size standards of less than 200 kb may be prepared by annealing P2 (30-kb monomer) or P4 (11.2-kb monomer) bacteriophage DNAs (20). *S. cerevisiae* chromosomal DNAs are useful as additional size markers (20). Molecules larger than 1 Mb may be sized by comparison with the three *S. pombe* chromosomes (Fig. 1) or the multiple chromosomes of other yeasts (S. Ringquist, J. B. Fan, and C. L. Smith, submitted for publication).

Circular bacterial chromosomal DNAs must be linearized before they can be sized directly by PFG (Fig. 1). This can be done by cleavage at a single site or by brief exposure to gamma irradiation, DNase I, or a restriction enzyme (38). For random cleavages a titration is done to maximize the fraction of chromosomes with a single double-strand break. By Poisson statistics, this will only be about one third of the molecules. Many DNA preparations contain enough randomly broken molecules so that no further breakage is necessary. The titration experiments will also confirm the presumed circular chromosome topology.

New cutting methods. A variety of proposed methods have the potential to cut small genomes once. However, DNA methylation-restriction enzyme combination strategies have not yet worked reliably on large DNA molecules (39). Chemical cleavage reactions (36) and methylation protection strategies (15) have not yet been tested with genomic DNAs. A variety of site-specific recombination enzymes may prove useful in the future. For instance, in vivo induction of bacteriophage λ terminase cuts a chromosomal λ *cos* site (Smith, unpublished observations), but this enzyme has not yet been tried in vitro. The use of bacterial chromosomal DNA substrates in restriction enzyme searches may uncover other enzymes undetected with small DNA substrates.

Detection of extrachromosomal elements. PFG of uncut genomic DNA with very short pulse times (0.5 to 10 s) will reveal the presence of extrachromosomal elements which may or may not be circular (13). Linear and circular DNAs respond differently to pulse times (12, 18, 19). For instance, the size of a circular DNA will appear to depend on pulse time if its mobility is compared with those of linear size standards. Hence, the topology of an extrachromosomal element may be determined by comparing its PFG mobility with those of linear or circular size standards at different pulse times. Alternatively, the circularity of DNA may be confirmed by converting it to a linear molecule or by PFG with fields that are 180° apart (FIGE; 7).

It is thought that relaxed circular DNAs larger than about 15 kb do not migrate in agarose because they get impaled on the agarose like a ring on a spear. Such DNAs will move in agarose when pulsed fields that are 180° apart are used. In FIGE, DNA molecules move forward and backward; net mobility requires uneven pulsing, e.g., a forward pulses longer than the reverse pulse. Such movement apparently frees entrapped circular molecules, which supports the impalement model (18). FIGE mobility is not a monotonic function of DNA size (9). Hence, FIGE should not be used for sizing DNA.

CHROMOSOMAL RESTRICTION MAP CONSTRUCTION AND GENE MAPPING

Random clones as hybridization probes. The original, low-resolution, *E. coli* K-12 restriction map was constructed by using genetically mapped, cloned genes as hybridization probes to PFG-fractionated complete and partial *Not*I digests of chromosomal DNA (31). Since randomly chosen probes are not likely to span rare restriction sites, they will hybridize to an unknown location on a single large restriction fragment in a total digest (Fig. 2A). In a partial digest such probes will hybridize to a series of fragments that extends in both directions from the total digestion fragment (Fig. 2C) unless the probe is from a chromosome end (Fig. 2D). In a circular genome, the latter might be generated by a very rare naturally occurring or engineered restriction site. In partial-digest experiments, the difference in band sizes gives the physical distance between the rare cutting sites. For small genomes, ten probes should allow enough overlap in the digestion pattern to permit the construction of a map.

Most microorganisms do not have a large number of genetically mapped cloned genes available. Instead, hybridization probes can be chosen randomly from a species-specific library or even from cloned genes of other organisms. For instance, the use of conserved sequences like rDNAs will reveal the number of genes and their chromosomal distribution (27).

DNA fragments as hybridization probes. A large fragment (or a clone from it) generated by one enzyme can be used as a hybridization probe to identify overlapping large fragments generated by a second enzyme (Fig. 2B; 17). Probing a partial genomic digest, generated with the same or a different enzyme, allows several adjacent fragments to be identified (Fig. 2C and D). This approach circumvents problems associated with clustering of rare cutting sites as in the *Haemophilus influenzae* Rd genome (17). Hybridization of repeated sequences (e.g., rDNA genes or insertion elements) to nonadjacent chromosomal fragments can be eliminated by preannealing the probe or by blocking DNA on a membrane with the specific sequence or with sheared chromosomal sequences.

Two-dimensional PFG was used to construct a physical map of the 1-Mb *Mycoplasma mobile* genome (3). Chromosomal DNA digested either partially or completely by one

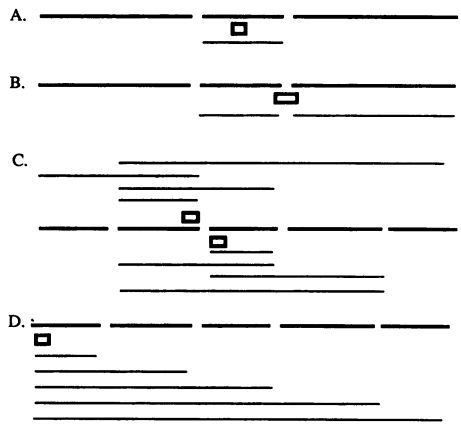

FIG. 2. Strategies for constructing low-resolution genomic restriction maps; probes (□), restriction fragments (——), and DNA bands (▬) are indicated. In a complete genomic digest (A) most probes will hybridize to a single large restriction fragment; linking probes (B) will hybridize to two adjacent large genomic restriction fragments. In partial digests (C) a probe will hybridize to a series of bands that extend bidirec tionally from the smallest fragment containing the homologous sequence; adjacent halves of a linking clone will detect some coincidental bands. Hybridization with probes from chromosome ends (D) will identify fragments extending unidirectionally.

restriction enzyme was fractionated in the first dimension with PFG. DNA in a thin slice or a slab of a gel lane was digested in situ to completion with the same enzyme or with a second enzyme, respectively, and a second dimensional analysis was done by PFG or conventional electrophoresis. Common double-digestion products signal overlap of genomic fragments generated by different enzymes, while complete-digest products identify the restriction fragment constituents of individual partial-digest fragments. Application of this approach to larger genomes containing more, and sometimes overlapping, fragments may require hybridization.

Linking clones as hybridization probes. Linking libraries are composed of clones having rare restriction enzyme sites. A linking clone used as a hybridization probe will identify two adjacent genomic fragments (Fig. 2B). Linking probes are particularly useful in the analysis of partial digests. Each half of a linking probe will identify a different set of restriction fragments. Several different restriction maps may be consistent with each set of data, but the true restriction map must be consistent with both sets (Fig. 2C).

Preparation of linking libraries is not yet routine (29, 34). One approach is to cut a circular, plasmid or λ, genomic library made with one enzyme with a second, infrequently cutting restriction enzyme and to ligate a selectable marker into those few clones that are cut. The linking clones can then be selected for after transformation into an appropriate host strain. Circular λ libraries are created by ligating the λ *cos* ends together intramolecularly, at a low DNA concentration. Linking-library approaches are less likely to miss small fragments since each rare restriction site has an equal probability of being cloned. In contrast, random approaches favor selection of clones from large fragments.

Protocols. Many hybridization procedures work successfully with PFG. The major trick appears to be the use of UV irradiation, rather than HCl treatment, for DNA nicking before transfer from the gel to a membrane (32). DNA fragments (or clones from fragments) used as hybridization probes can be labeled by using the oligonucleotide-primed labeling procedure directly in agarose, although the specific activity may be slightly lower. High-specific-activity probes are needed to compensate for the small amounts of DNA that are loaded and recovered from agarose. Fragments may be recovered from agarose by various standard procedures, such as agarase treatment, electroelution by PFG, or standard electrophoresis. Alternatively, DNA bands may be cut out of agarose and freeze-thawed several times in the presence of phenol, and then the DNA may be recovered from the supernatant fraction after centrifugation. These procedures will shear the fragments, unless special precautions are taken. For many applications, including labeling, this will not matter. However, libraries are most conveniently made from large fragments by in situ restriction enzyme digestion in agarose to prevent shear damage. DNA fragments needed for any subsequent purposes should be exposed to the least possible amount of UV irradiation during gel photography (6).

Physical mapping of transposon insertions. Transposon Tn*5* (in its IS*50* sequences) and plasmid F^+ both have *Not*I sites (33). Thus, the location of these genetic elements in the chromosome is marked by *Not*I sites, and these elements are linking clones for strains which contain them. Their approximate chromosomal location is revealed by comparing the ethidium bromide-stained DNA fragment patterns of isogenic strains with and without the element. The sum of the two new *Not*I fragments should equal the size of the parental fragment. The precise localization of the Tn*5* insertion requires that the polarity of the two new fragments be determined by using a clone from one end of the parental fragment as a hybridization probe.

Use of transposons in restriction map construction. Many organisms without well-developed genetic systems have biologically interesting genes mutated via Tn*5* insertions or linked to nearby Tn*5* insertions. These insertions can be used to construct physical maps and to align them with genetic maps. When Tn*5* is hybridized to DNA partially digested with *Not*I, a series of fragments will be detected that extends bidirectionally from the Tn*5* insertion. Of course, Tn*5* is not useful for all organisms. Some GC-rich genomes may already have too many *Not*I restriction sites. The transposon IS*5*, which has an *Avr*II site (Condemine and Smith, in press), is potentially useful in some cases, while in others, construction of appropriate transposons may be required.

Recombination of rare restriction sites into the chromosome. Cloned genes can be genetically mapped by introducing selectable markers into the homologous chromosome region by recombination. Chromosomal integration of circular plasmids is promoted by growing cells in low-phosphate medium (25), by plasmid incompatibility displacement of a resident plasmid (26), by transforming a plasmid into *polA* cells (11), or by transforming a linear fragment or plasmid into *recBC sbcB* cells (2).

Recombination can mark a gene location for PFG detection. Integration of a circular plasmid into the chromosome results in a gene duplication that increases the size of the restriction fragment containing it. In *E. coli*, precise gene mapping can be done by using a plasmid containing a *Not*I restriction site (30). This introduces a new *Not*I site at or, in the case of an essential gene, nearby a gene. Precise mapping requires determining the polarity of the two new *Not*I fragments as described above. These experiments will determine the direction of transcription on the chromosome if the direction on the plasmid is known.

Identification of the DNA replication origin and terminus. Simple PFG experiments can identify the chromosomal DNA replication origin and terminus and track a replication fork moving along the chromosome in synchronously growing cells (22). Replication can be synchronized by inhibiting protein synthesis since initiation but not elongation requires de novo protein synthesis. When the block in protein synthesis is removed, DNA synthesis initiates from the replication origin and proceeds synchronously around the chromosome for at least one round of DNA replication. Large restriction fragments containing the synchronized replication forks can be identified by pulse-labeling with [^{14}C] thymidine. The first fragment labeled after reinitiation contains the origin, whereas the last fragment labeled after shutdown or reinitiation contains the terminus. Such experiments can also tentatively order the fragments between the origin and the terminus.

COMBINING COMPUTATIONAL AND PHYSICAL MAPPING APPROACHES

DNA data bases. About 17% of the *E. coli* chromosome is available for analysis through the DNA sequence data bases. Faster DNA-sequencing strategies may soon provide the missing *E. coli* sequences. Genes then can be localized by sequence alignment since only a small amount of sequence (<20 base pairs) should serve as a unique fingerprint for most chromosomal loci. In some cases, physical methods will still be required to confirm chromosomal location.

High-resolution restriction maps. High-resolution genomic restriction maps, like that of Kohara et al. (14), can also be used to assign gene location. For instance, a cloned gene may be located on a *Not*I fragment by hybridization or genetically mapped to a region. The fine restriction map of the clone of interest then can be aligned with a high-resolution genomic restriction map by a simple manual method (1) or by computational methods eventually. Alternatively, the clone of interest can be hybridized to a subset of an overlapping genomic library (14, 37) which covers the identified region. Cloned genes can also be mapped by hybridization to an entire library directly. However, routine scanning of least several hundred clones is not yet available.

INTEGRATED APPROACHES TO FINE GENOMIC MAPPING

Top-down genomic sequencing strategies. Various integrated approaches may be used to obtain genomic restriction maps, overlapping libraries, and, eventually, complete genomic sequences. The easiest task, the construction of low-resolution restriction maps by using PFG, can serve as a basis for top-down integrated genomic mapping and sequencing projects. In contrast, the bottom-up restriction mapping (14) and sequencing (8) approaches currently are technically difficult. However, these latter approaches are assisted by large-DNA technology.

High-resolution restriction maps. High-resolution genomic restriction maps can be made by hybridizing half of a linking clone to genomic DNA totally digested with the corresponding rare cutter and partially digested with a restriction enzyme which cuts frequently. In such experiments fragments are identified that extend in one direction from the rare cutting site. The best PFG experiments will resolve fragments that differ by 1% in size. Restriction site clusters closer than this are indistinguishable, so construction of finer maps will require an overlapping library.

Overlapping genomic libraries. Kohara et al. (14) constructed a complete overlapping λ library of *E. coli* by fingerprinting clones with eight restriction enzymes. Recently, Tabaka et al. (37) constructed a partially complete *E. coli* cosmid library by integrating several physical and genetic approaches. Specific clones were identified by hybridization to known genes and were used as probes in chromosome-walking experiments to identify sets of contiguous overlapping clones called contigs. Correlation of the contigs with specific chromosomal regions was confirmed by hybridizations to mapped genes and to *Not*I restriction fragments and by genetic complementation of mutations. Wenzel and Herrman (40) have constructed a complete overlapping cosmid library of the small *Mycoplasma pneumoniae* genome by using a fast chromosome-walking vector to expand contigs.

Large-DNA techniques can increase the efficiency of clone sorting. For instance, large fragments can be used as hybridization probes to sort clones. Clone banks can be made directly from large fragments. Large fragments can also be used to span chromosomal regions containing unconnected contigs.

Complete sequencing of small genomes. Automatic sequencing devices allow massive amounts of DNA sequence to be generated daily, provided DNA samples are available. Automated DNA preparation should eventually be possible. Church (8) is using a multiplex strategy to sequence the *E. coli* and *Salmonella typhimurium* chromosomes. In this strategy, chromosome libraries are constructed, simultaneously, in multiple vectors. A multiplex library is constructed by pooling a clone from each of the libraries. Sequencing reactions are performed with each pool; the products are fractionated on a DNA-sequencing gel and then transferred to a nylon membrane. Since each vector has unique sequences on each side of the cloning site, blots, containing different pools, can be hybridized en masse to an oligonucleotide specific for one of these sequences. The set of blots can be successively reprobed with oligonucleotides recognizing different unique vector sequences. Hence, labor is conserved in the sequencing reactions, in gel running and blotting, and in hybridization.

The identification of the most efficient sequencing approach awaits evaluation of experimental results. At present, a genomic-sequencing project for any microorganism requires a major long-term commitment, while in the near future the effort involved may be substantially reduced. Thus, such efforts seem a bit premature unless they are used to develop new technologies or are needed to answer a specific biological question.

PROSPECTUS

Megabase polymorphisms detected in single PFG experiments revealed relationships between *Legionella* (A. Marra, H. Shuman, and C. L. Smith, unpublished results), *Halobacterium* (29), *Mycobacterium* (W. Jacobs, B. Bloom, and C. L. Smith, unpublished results) and *E. coli* (22, 33) isolates. However, technology needs to be developed to simplify map construction by using maps that already exist for related organisms.

ACKNOWLEDGMENTS

We thank H. Shuman, S. Chang, R. Amils, A. Lichens-Park, R. Aasland, and C. R. Cantor for discussions on the applications and use of the new technologies.

This work was supported by the Department of Energy (DG-FG-02-87ER-GD825) and the MacArthur Foundation.

LITERATURE CITED

1. **Aasland, R., and C. L. Smith.** 1988. An efficient manual method for aligning DNA restriction map data on very large genomic restriction maps. Nucleic Acids Res. **16**:10392.
2. **Bassett, C. L., and S. R. Kushner.** 1984. Exonucleases I, III, and V are required for stability of ColE1-related plasmids in *Escherichia coli*. J. Bacteriol. **164**:51–56.
3. **Bautsch, W.** 1988. Rapid physical mapping of the *Mycoplasma mobile* genome by two-dimensional field inversion gel electrophoresis techniques. Nucleic Acids Res. **16**:11461–11467.
4. **Cantor, C. R., A. Gaal, and C. L. Smith.** 1988. High resolution separation and accurate size determination in pulsed field gel electrophoresis. III. Effect of electrical field shape. Biochemistry **27**:9216–9221.
5. **Cantor, C. R., C. L. Smith, and M. K. Mathew.** 1987. Pulsed field gel electrophoresis of very large DNA molecules. Annu. Rev. Biophys. Biophys. Chem. **17**:287–304.
6. **Cariello, N. F., P. Koehavong, B. J. S. Sandrson, and W. G. Thilly.** 1988. DNA damage produced by ethidium bromide staining and exposure to ultraviolet light. Nucleic Acids Res. **16**:4157.
7. **Carle, G. F., M. Frank, and M. V. Olson.** 1986. Electrophoretic separations of large DNA molecules by periodic inversion of the electric field. Science **232**:65–68.
8. **Church, G., and S. Kieffer-Higgins.** 1988. Multiplex DNA sequencing. Science **249**:185–188.

8a. **Condemine, G., and C. L. Smith.** 1990. Genetic mapping using large-DNA technology: alignment of *Sfi*I and *Avr*II sites with the *Not*I genomic restriction map of *Escherichia coli* K-12, p. 53–60. *In* M. Riley and K. Drlica (ed.), The bacterial chromosome. American Society for Microbiology, Washington, D.C.

9. **Ellis, T. H. N., W. G. Cleary, K. W. G. Burcham, and B. A. Bowen.** 1987. Ramped field inversion gel electrophoresis: a cautionary note. Nucleic Acids Res. **15**:5489.
10. **Fan, J.-B., Y. Chikashige, C. L. Smith, O. Niwa, M. Yanagida, and C. R. Cantor.** 1989. Construction of a *Not*I restriction map of the fission yeast *Schizosaccharomyces pombe* genome. Nucleic Acids Res. **17**:2801–2818.
11. **Greener, A., and C. W. Hill.** 1980. Identification of a novel genetic element in *Escherichia coli* K-12. J. Bacteriol. **144**: 312–321.
12. **Hightower, R. C., D. W. Metge, and D. V. Santi.** 1987. Plasmid migration using orthogonal-field-alternation gel electrophoresis. Nucleic Acids Res. **15**:8387–8398.
13. **Kinashi, H., M. Shimaji, and A. Sakai.** 1987. Giant linear plasmids in Streptomyces which code for antibiotic biosynthesis genes. Nature (London) **328**:454–456.
14. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50**:495–508.
15. **Koob, M., E. Grimes, and W. Szybalski.** 1988. Conferring operator specificity on restriction endonucleases. Science **241**: 1084–1086.
16. **Lee, J. J., and H. O. Smith.** 1988. Sizing of the *Haemophilus influenzae* Rd genome by pulsed-field gel electrophoresis. J. Bacteriol. **170**:4402–4405.
17. **Lee, J. J., H. O. Smith, and R. J. Redfield.** 1989. Organization of the *Haemophilis influenzae* Rd genome. J. Bacteriol. **171**: 3016–3024.
18. **Levene, S. D., and B. H. Zimm.** 1987. Separations of open-circular DNA using pulsed-field electrophoresis. Proc. Natl. Acad. Sci. USA **84**:4054–4057.
19. **Mathew, M. K., C.-F. Hui, C. L. Smith, and C. R. Cantor.** 1988. High resolution separation and accurate size determination in pulsed field gel electrophoresis of DNA. IV. The influence of DNA topology. Biochemistry **27**:9222–9226.
20. **Mathew, M. K., C. L. Smith, and C. R. Cantor.** 1988. High resolution separation and accurate size determination in pulsed field gel electrophoresis of DNA. I. DNA size standards and the effect of agarose and temperature. Biochemistry **27**:9204–9210.
21. **Mathew, M. K., C. L. Smith, and C. R. Cantor.** 1988. High resolution separation and accurate size determination in pulsed field gel electrophoresis of DNA. II. Effect of pulse time and electric field strength, and implications for models of the separation process. Biochemistry **27**:9211–9216.
22. **Ohki, M., and C. L. Smith.** 1989. Tracking bacterial replication forks *in vivo* by pulsed field gel electrophoresis. Nucleic Acids Res. **17**:3479–3490.
23. **Orbach, M. J., D. Vollrath, R. W. Davis, and C. Yanofsky.** 1988. An electrophoretic karyotype of *Neurospora crassa*. Mol. Cell. Biol. **8**:1469–1473.
24. **Pyle, L. E., L. N. Corcoran, B. G. Cocks, A. D. Bergemann, J. C. Whiteley, and L. R. Finch.** 1988. Pulsed field gel electrophoresis indicates larger than expected sizes for Mycoplasma genomes. Nucleic Acids Res. **16**:6015–6025.
25. **Roeder, D. L., and A. Collmer.** 1985. Marker-exchange mutagenesis of a pectate lyase isozyme gene in *Erwinia chrysanthemi*. J. Bacteriol. **164**:51–56.
26. **Ruvkun, G. B., V. Sundaresan, and F. M. Ausubel.** 1982. Directed transposon Tn*5* mutagenesis and complementation analysis of *Rhizobium meliloti* symbiotic nitrogen fixation genes. Cell **29**:551–559.
27. **Sanz, J. L., I. Marin, L. Ramirez, J. P. Abad, C. L. Smith, and R. Amils.** 1988. Variable rRNA gene copies in extreme Halobacteria. Nucleic Acids Res. **16**:7827–7832.
28. **Schwartz, D., W. Saffran, J. Welsh, R. Haas, M. Goldenberg, and C. Cantor.** 1983. New techniques for purifying large DNA's and studying their properties and packaging. Cold Spring Harbor Symp. Quant. Biol. **47**:189–195.

28a. **Smith, C. L., J. P. Abad, and G. Condemine.** 1988. Pulsed field gel electrophoretic snapshots of chromosomes, p. 152–157. *In* K. Okuda (ed.), The Congress Proceedings of the 3rd ANT. Blackwell Scientific Publishers, Oxford, England.

29. **Smith, C. L., and C. R. Cantor.** 1987. Purification, specific fragmentation, and separation of large DNA molecules. Methods Enzymol. **155**:449–467.
30. **Smith, C. L., G. Condemine, S.-Y. Chang, E. McGary, and S. Chang.** 1989. Insertion of rare cutting sites nearby genes allows their rapid physical mapping: localization of the *E. coli map* gene. Nucleic Acids Res. **17**:817.
31. **Smith, C. L., J. Econome, A. Schutt, S. Klco, and C. R. Cantor.** 1987. A physical map of the *E. coli* K12 genome. Science **236**:4481–4490.
32. **Smith, C. L., S. Klco, and C. R. Cantor.** 1988. Pulsed field gel electrophoresis and the technology of large DNA molecules, p. 41–72. *In* K. Davies (ed.), Genome analysis: a practical approach. IRL Press, Oxford, England.
33. **Smith, C. L., and R. D. Kolodner.** 1988. Physical mapping of *E. coli* chromosomal Tn*5* and F insertions by PFG electrophoresis. Genetics **119**:227–236.
34. **Smith, C. L., S. K. Lawrance, G. A. Gillespie, C. R. Cantor, S. M. Weissman, and F. S. Collins.** 1987. Strategies for mapping and cloning macro-regions of mammalian genomes. Methods Enzymol. **151**:449–467.
35. **Smith, C. L., T. Matsumoto, O. Niwa, J. B. Fan, M. Yanagida, and C. R. Cantor.** 1987. An electrophoretic karyotype of *Schizosaccharomyces pombe* by pulsed field gel electrophoresis. Nucleic Acids Res. **15**:4481–4490.
36. **Strobel, S. A., H. E. Moser, and P. Devan.** 1988. Double strand cleavage of genomic DNA at a single site by triple helix formation. J. Am. Chem. Soc. **110**:7927–7929.
37. **Tabaka, S., A. Hlgashitani, M. Takanami, K. Akiyama, Y. Kohara, Y. Nishimura, A. Nishimura, S. Yasuda, and Y. Hirota.** 1989. Construction of an ordered cosmid collection of the *Escherichia coli* K-12 W3110 chromosome. J. Bacteriol. **171**: 1214–1218.
38. **Van der Bliek, A., C. R. Lincke, and P. Borst.** 1988. Circular DNA of 3T6R50 double minute chromosomes. Nucleic Acids Res. **16**:4841–4851.
39. **Weil, M. D., and M. McClelland.** 1989. Enzymatic cleavage of a bacterial genome at a 10-base-pair recognition site. Proc. Natl. Acad. Sci. USA **86**:51–55.
40. **Wenzel, R., and R. Herrman.** 1988. Physical mapping of the *Mycoplasma pneumoniae* genome. Nucleic Acids Res. **16**: 8323–8336.