

Physical Mapping of Bacterial Genomes

DONNA L. DANIELS

138

INTRODUCTION

The use of physical maps for the description and comparison of bacterial genomes has dramatically increased in the last few years with the advent of powerful tools for the construction of physical maps of large DNA molecules. The availability of restriction enzymes that cut whole bacterial genomes only a few times together with the use of pulsed-field gel electrophoresis to resolve and size the resulting fragments has allowed the direct restriction mapping of bacterial genomes. Nearly 50 bacterial genomes have been mapped for one or more of these enzymes. The isolation of complete ordered clone banks (encyclopedias) has been achieved for *Escherichia coli*, *Rhodobacter capsulatus*, *Mycoplasma pneumoniae*, *Mycobacterium leprae*, *Halobacterium* sp. and *Haloferax volcanii* PS2. By merging maps of the clones, high-resolution restriction maps of complete bacterial genomes have been obtained. Also, the accessibility of sequence databases, software tools, and curated databases for particular genomes has improved the usefulness and expanded the utility of both the maps and the encyclopedias.

In this paper I review (i) the low-resolution maps of *E. coli* and other bacterial genomes obtained directly by fragmentation and measurement of fragments on pulsed-field gels of whole bacterial DNA, (ii) the bacterial genome encyclopedias and methods for isolating these complete ordered clone banks, and (iii) uses to which the maps and the encyclopedias can be put.

PHYSICAL MAPS

A map is an enumeration of the locations of sites on an object. In molecular biology, the sites are genes, protein-binding sites, restriction sites, sequence-tagged sites (STSs), etc., and the object is a DNA molecule. Our maps are one dimensional. Maps are presented either graphically as a scale drawing or tabularly as a list of sites and their coordinates. The information contained in a map is the relative order of all the sites and the distance between every pair of sites on the map. For n sites there are $n(n-1)/2$ intervals whose lengths can be obtained by subtracting the coordinate of one site from that of another. A map is constructed by measuring the intervals between a few of the pairs of sites and deducing the map from these measurements; thus, a map contains much more information than a mere tabulation of the input measurements. Minimally, the data needed to deduce a map must include for every site one interval ending at that site, and the entire set of sites needs to be connected via some data path of measured intervals. A physical map is one in which the measurements used to deduce the map are of the physical distances between sites on the DNA molecule (as opposed to genetic recombination frequencies). Distances are additive.

Restriction maps are a common type of physical map. The sites on the map are restriction enzyme cut sites, and distances between the restriction sites are obtained by measuring the sizes of DNA fragments that result from digestion with the restriction enzymes. Typically, fragment sizes are measured by determining their mobilities in electrophoretic gels and comparing these mobilities to the mobilities of standards of known length. Although a physical map has come to be almost synonymous with a restriction map, other physical maps have been constructed (93). For example, the detailed physical map of lambda DNA that enumerates the regions of similarity between lambda and other lamdoid phages and of lamdoid hybrid recombination points was determined from electron microscopic measurements of DNA intervals between

bifurcation points in heteroduplexes (34). For mammalian genomes, STSs have become preferred sites for physical mapping (71). However, no good experimental method for measuring the physical distances between ordered STSs is available. Thus, mapping them has depended on ordering these sites relative to other mapped entities such as clone end points.

RESTRICTION MAPPING

Several techniques can be used for constructing restriction maps of a DNA molecule from fragmentation data. The classic method is to completely digest the DNA in three ways: with one restriction enzyme, with a second enzyme, and with a combination of both. The resulting fragments are counted and measured. Because they are physical measurements, the lengths of all intervals and subintervals on the map are additive (within measurement error). The fragmentation pattern, the fragment sizes, and the requirement for additivity are used to deduce the order of sites and the assignment of measured fragments to intervals between sites. This analysis is combinatorially complex, and the complexity increases rapidly with increasing numbers of cut sites per enzyme and of enzymes. Once a map is deduced (site order, fragment assignment, interval measurement), the map of the molecule can be refined numerically by choosing coordinates for the sites so as to minimize the sum of squares of the fractional deviation between intervals as calculated by subtraction of coordinates and as measured for fragmentation data (86).

Technically, the digestions and fragment measurements are simple to perform, but the method has a number of limitations. First, it is important to detect all of the fragments resulting from digestion of the molecule to be mapped. Failure to detect small fragments is common, because they run off the bottom of the gel or are not visible by staining because of their low mass. It is frequently not possible to load enough DNA to detect the smaller fragments. Fragments can also be missed because they are not resolved from similarly sized fragments. These bands are often called “doublets” and are sometimes detectable because they stain darker than would be expected from the intensity of similarly sized fragments. The necessity to resolve all the fragments limits the number of fragments that can be mapped and thus limits the size of the DNA molecule to be mapped, depending on the frequency of cutting of the enzyme. During analysis, the existence of an undetected fragment can often be deduced by inconsistencies in the number of fragments in the three digests. Second, since the analysis depends on additivity, it is important to get reasonably accurate sizes for all of the fragments (say, within 5%). This level of accuracy requires the use of accurate size standards in the appropriate size range on gels that resolve fragments in the necessary size range. It is frequently necessary to use more than one gel system and more than one standard in order to cover the entire size range of the fragments to be resolved and measured. The availability of size standards limits the sizes of genomes and fragments to be mapped. Finally, and most fundamentally, the method is limited by the combinatorial complexity of the analysis problem. If all the fragments are detected and there is no measurement error, the size of the configuration space of possible arrangements of sites and fragments is $A!B!(A + B - 1)!$, where A is the number of fragments for enzyme A, B is the number of fragments for enzyme B, and $A + B - 1$ is the number of fragments in the double digest (if the molecule is linear). If one allows measurement error (as in real problems), then the order of the sites is not unambiguously determined by the order of the fragments and the size of the problem is $A!B!(A + B - 1)!(A + B - 2)!$, where $A + B - 2$ is the number of sites. For practical purposes, this limits the method to maps with perhaps six or seven sites per enzyme. Maps with many restriction sites can be derived stepwise if there are only a few sites for each enzyme. Analysis starts with the “easy” enzymes (few sites of distinctively different sizes) and proceeds one enzyme (a few sites) at a time to denser maps. The prior constraints of an existing map ease the combinatorial difficulties and aid in the choice of which experimental data to gather to give definitive information about additional enzymes to be mapped.

Smith-Birnstiel mapping (90) reduces the computational complexity but introduces significant other technical problems. In this method, the DNA to be mapped is labeled at a unique end (site), the sample is divided, and each aliquot is partially digested with an enzyme to be mapped. The set of partial digests is

run in adjacent lanes in an electrophoresis gel. Autoradiography reveals a conceptually easy-to-decipher pattern, with each lane containing bands of fragments that start at the unique site and end at a site to be mapped. The order of sites can be read up the gel, from the smallest fragment (site closest to the labeled site) to the largest (farthest from the site). The distance between adjacent sites is the difference in the sizes of the fragments. The lengths of all the fragments in all the lanes constitute the set of all intervals from the unique labeled site to all sites on the map. A major problem with such maps is that the accuracy of the map decreases with the distance of the mapped sites from the labeled site because of the decreasing ability of a single gel to resolve larger and larger fragments. Similarly, the span over which a map can be read depends on the size range that can be resolved on the gel. Ideally, the gel would need to resolve and accurately measure sizes from the smallest fragment (labeled site to closest enzyme cut site) all the way to full-length undigested molecules. Another difficulty is finding a unique site to label. End labeling with kinase followed by digestion with another enzyme to remove the other end and yield an asymmetrically end-labeled fragment to be mapped requires significant work and previous knowledge of the map. Indirect labeling by hybridization of labeled probe to a site near an end is more straightforward but requires a clone or synthesis of a unique region near a unique end (32, 80, 89). Finally, it is experimentally difficult to obtain appropriate partial digests for several different enzymes. Differences between various lots of enzyme and different DNA preparations make it almost a case-by-case problem. Further, within a DNA molecule, different sites have different rates of cutting, so the method, which depends on partial digests being representative of the map, may miss a site. This can be somewhat alleviated by methods that rely on partial methylation (41).

STRATEGIES FOR MAPPING BACTERIAL GENOMES

The discussions above refer to methods of making complete maps of an entire molecule directly by measuring fragments of the DNA molecule itself. A map is complete if all sites for that enzyme are mapped on the specified DNA. Mapping based on size additivity with no information other than fragmentation pattern is limited to a handful of sites for any given enzyme because of the computational complexity and technical limitations. For a large molecule such as an entire bacterial genome, the constraints are prohibitive. The few enzymes available that cut only rarely generally yield more than the handful of fragments to be mapped, the fragments are large and have to be resolved on pulsed gels (87), the accuracy of the measurements is problematic for fragments larger than perhaps 1 Mb, and detecting fragments smaller than perhaps 15 kb often requires special procedures such as end labeling. Therefore, to map more than a handful of sites, additional information about relationships between fragments is experimentally obtained. The relationship can be fragment adjacency or overlap. These relationships are detected when two fragments from the same (complete) digest both hybridize with a probe that has a site in it (a linking probe) or when two fragments from different digests hybridize to the same probe. Information about fragment relationships together with fragment size measurements and the requirement for additivity has been used to derive complete maps of a number of bacterial genomes for many enzymes. Restriction maps of bacterial genomes obtained in this way are generally termed complete, low-resolution maps (fragment sizes averaging 100 to 500 kb). Fragment length measurement requires the use of pulsed-field gels (87).

Determining Fragment Overlaps

The many methods for determining fragment relationships might be categorized as follows. (i) Hybridize genomic clones of the DNA to be mapped to genomic Southern blots of pulsed-field mapping gels of single and double digests for the enzymes to be mapped. (ii) Isolate fragments to be mapped from mapping gels, and hybridize each of them to an array of dot blotted clones. (iii) Isolate fragments to be mapped, and hybridize them to a Southern blot of mapping gels. (iv) Compare restriction patterns between closely related genomes that have a defined genetic difference between them. (v) Search sequence data banks for

restriction sites. (vi) Look for fragment-inclusive relationships by size additivity in partial digests or for contained subfragments in isolated and redigested fragments (either individually or on two-dimensional gels).

Hybridizing genomic clones to blots of gels of genomes digested with two enzymes singly and in combination identifies fragment ordering. Two fragments in different digests hybridizing to the same probe indicates an overlap relationship between the fragments. Two fragments within the same digest hybridizing to the same probe indicates an adjacency relationship. Such probes are called linking probes and contain a restriction site within them. Methods to specifically clone linking probes have been described elsewhere (62, 77). Use of many probes together with the requirement for additivity can yield a complete map. Clones for probes can be anonymous or specific. If the clones have been previously mapped (e.g., to the genetic map), then fragments with which they hybridize are located to the genetic map (3). Thus, even in the absence of coincident fragment hybridization, the relative order of fragments can be determined, since the order of the probes is known. Hybridization of genetically mapped *E. coli* DNA clones to Southern blots of gels of digested DNA was used extensively in the determination of the first complete *E. coli* restriction map, that of *NotI* (88). Because the map locations of the probes were known, it was possible to order hybridizing fragments to the genetic map. The main problem with using this method exclusively is that small fragments are difficult to locate, since the chances of finding a clone that hybridizes is proportional to the fragment length. Technically, the use of very large numbers of clones becomes a lot of work, since each needs to be prepared, labeled, and hybridized.

The same information can be gathered by doing the experiment the other way. Fragments can be isolated from pulsed-field gels and used as probes against an array of dot blots of clones, either anonymous or mapped (e.g., the entire encyclopedia). Membranes can be prepared either by lifting “spots” of clones from plates or by dot blotting directly onto membranes. In either case, replicators, multichannel pipettes, or robotic pipetters can be used. A large number of clones and a few fragments are logistically easier to deal with, since clones do not have to be handled individually. This method was used to make an *AvrII* map of *E. coli* (24). Technically, the method may be limited to probes smaller than about 600 kb owing to background hybridization to host DNA in the dot blots.

Fragments can be isolated from pulsed-field gels and used as probes against Southern blots of mapping gels to determine overlapping fragments. This method was used extensively in the *E. coli BlnI-XbaI* map (76).

Comparison of fragmentation patterns of genomes whose genetic relationship is known is another method of ordering fragments on the genome (88). If the restriction patterns of two genomes differ only by fragment X in digest A and only by fragment Y in digest B, then X and Y overlap. Further, if the genetic map position of the difference is known, then fragments X and Y are located in relation to the genetic map as well as to each other. An important example of this type of mapping is the construction and restriction fragmentation of a set of transposon-inserted genomes (97). The resulting difference in restriction pattern is easier to detect by using a transposon with sites for many of the enzymes to be mapped. In comparing the original genome and the transposon-inserted genome, one fragment in each digest will be converted to two smaller fragments. This locates fragments in each digest with respect to each other. A large set of transposons can be isolated genetically and chosen randomly or selected specifically. For example, Weinstock (98) used a set of auxotrophic transposon insertions. These can be genetically mapped to correlate the physical and genetic maps.

For *E. coli*, with its wealth of information, other tools are available (8, 60, 83–85). The sequence data can be scanned for restriction sites, large numbers of strains with known chromosomal differences are available, and several complete clone banks are available.

Composite Maps

One can also construct maps indirectly. Maps of isolated subregions of the target genome can be constructed (by mapping clones) and extrapolated to the original genome. More extensive maps can be determined by merging of the subregion maps. To completely map a genome this way requires the isolation of a complete ordered set of clones spanning the genome. Since the clones are the mapping substrates, the experimental limitations apply to the clones. Fragments to be measured are in a more tractable size range, and concentration limitations are not a problem. Maps in which the average fragment is <20 kb long are called high-resolution maps. Frequently, composite maps are not complete over the entire genome for all the enzymes presented, and the accuracy varies over the length of the map (48). A particularly useful composite map containing not only restriction sites but also other types of physical mapping information has been compiled (83). This contains a complete (composite) restriction map constructed by merging the composite restriction map of Kohara et al. (48) with restriction maps predicted from sequence data, locations of sequence block end points, locations of sequenced-gene end points, and location of mapped clone end points.

Bacterial Genome Maps

Maps of nearly 50 bacterial genomes are available. These maps include those for *E. coli* (3, 10, 24, 25, 42, 47, 60, 75, 76, 88), *Salmonella typhimurium* (official designation, *Salmonella enterica* serovar Typhimurium) (55–57, 101), *Bacillus subtilis* (1, 43, 97), *Pseudomonas aeruginosa* (33, 40, 81–83), *Haemophilus* species (44, 53), *Mycoplasma* species (6, 21, 78, 79, 99), *Streptomyces* species (45, 51), *Caulobacter crescentus* (27, 29, 30), *Thermococcus celer* (67), *Bacillus cereus* (13, 49), *Bacillus thuringiensis* (14), cyanobacteria (4, 18), *Enterococcus faecalis* (63), *Serpulina hyodysenteriae* (105), *Brucella melitensis* (61), *Thermus thermophilus* (11), *Helicobacter pylori* (95), *Campylobacter jejuni* (9), *Bradyrhizobium japonicum* (50), *Haloferax* species (58, 96), *Neisseria meningitidis* (7), *M. leprae* (28), *Borrelia burgdorferi* (15, 69), *Rhodobacter* species (36, 92), *Listeria monocytogenes* (62), *Shigella flexneri* (68), and others.

Physical mapping of a bacterial genome is particularly attractive for those organisms for which extensive genetic tools are not readily available. A summary of the particular mapping methods used in achieving each map is given by Cole and Saint Girons (20).

GENE ENCYCLOPEDIAS

We use the term “encyclopedia” to refer to a complete ordered set of clones that spans the genome, since a random collection of clones is commonly termed a “library.” Kohara et al. (47, 48) reported an encyclopedia for *E. coli* W3110 in lambda vectors that had only eight small gaps. Birkenhohl and Vielmetter (10) reported an encyclopedia of *E. coli* BHB2600 in cosmid vectors that covered 95% of the genome with 12 gaps. We isolated an encyclopedia of *E. coli* MG1655 in lambda vectors that had seven gaps (25, 26; D. L. Daniels et al., submitted for publication). Knott et al. (46) reported a set of *E. coli* 803 clones in cosmid vectors in about 50 groups of overlapping clones. Tabata et al. (94) reported a set of *E. coli* W3110 clones in cosmid vectors in about 30 groups that cover about 70% of the genetic map.

Several other bacterial genomes have also been dissected into ordered clone encyclopedias. These include the genomes of *M. pneumoniae* (99, 100), *R. capsulatus* SB1003 (36), *Haloferax volcanii* DS2 (17), *M. leprae* (28), and *Halobacterium* sp. strain GRB (91).

Strategies for Constructing Encyclopedias

Strategies for isolating complete sets of clones can be categorized as random (bottom up) or directed (top down). In directed strategies, one specifically clones all designated pieces of DNA. This strategy tends to be very laborious and is prohibitively so when the goal is a complete set covering a large region.

Nonetheless, the *M. pneumoniae* (99, 100) encyclopedia was isolated by using a directed strategy. The technique involved cloning RI fragments into a cosmid vector having SP6 and T7 promoters and using riboprobes from both ends of the insert to “walk” around the 800-kb genome, thus isolating a stepwise series of adjacent clones from a library.

Random strategies rely on arbitrarily choosing a large number of clones from libraries and doing experiments to identify overlap relationships between pairs of clones. Eventually (at a severalfold excess of the number of clones that would minimally be required to span the project once), the overlap relationships and the redundancy lead to an ordering of the whole set.

Experimentally, there are many ways to detect clone overlaps, including hybridization of one clone or clone subpart to another, coincident hybridization of two clones with some other probe(s), map (or restriction site pattern) similarities, and sequence overlap (59). The entire data set can be thought of as an $n \times n$ matrix of zeros and ones (true and false), indicating for each pair of clones overlap or nonoverlap. The data could also have values such as not tested or ambiguous, or one could define a 0 to 1 metric of overlap probability. In actual data, there are false negatives, false positives, false adjacencies (due to cloning artifacts), and experimental gaps and ambiguities. Ordering of all the clones into a map from the data matrix is a nontrivial computational problem. The amount of data needed to choose a genome-spanning set of clones is a small subset of the entire $n \times n$ matrix, and directed approaches can be thought of as strategies for obtaining just the data needed for identifying a spanning set. Even in a random approach, it is not necessary to obtain data for all possible pairs of clones. For accurate data, every clone needs to be identified as overlapping something, and enough clones need to overlap more than one thing for the entire set to be connected.

Fingerprint Strategies

Since for n clones there are $n(n-1)/2$ pairwise combinations (over 10^6 for $<2,000$ clones), experimental techniques that deal individually with all possible pairs (such as by hybridization of every clone to every other) could be prohibitively laborious. Therefore, techniques that aim at identifying all pairs of overlaps between all pairs of clones on the basis of fingerprints have been developed. A fingerprint is defined as a characteristic set of features of a piece of DNA. Thus, instead of doing $n(n-1)/2$ experiments, one experiment is done on each clone to derive a fingerprint specific for the DNA of the clone (n experiments), and the $n(n-1)/2$ comparisons are done by computer. Overlaps are inferred between clones that have sufficiently similar fingerprints. A great advantage of this approach is that one does not have to commit at the beginning of the data gathering to an n (number of randomly chosen clones) to be analyzed. Clone ordering can be attempted as experimental data gathering proceeds, and data gathering can stop when ordering is successful. The stepwise process increases n , the number of clones, adding one row and one column to the overlap data matrix. The labor involved in increasing n is linear, not exponential.

Factors contributing to a good fingerprinting method are (i) sensitivity (can it identify overlaps between two clones even if the overlap is small and yet not predict false overlaps?), (ii) experimental ease in obtaining the fingerprint for a very large number of clones, and (iii) inherent usefulness of the fingerprint for other purposes.

Sensitivity could be viewed as the average percent overlap between clones that is needed to detect an overlap. It depends on how distinctive the components of the fingerprint are and how frequently they occur. I emphasize that the occurrence and distribution of fingerprint features are statistical processes. Conclusions about the overlap or nonoverlap of any particular pair of clones are not absolutes but, rather,

probabilities.

Lander and Waterman (51) derived formulae that describe the statistical expectations of progress on a project to overlap random clones. Progress is visualized by plotting number of gaps remaining versus average depth of coverage (sum of the lengths of all fingerprinted clones divided by the genome length). The profile shows that the project is very fragmented until close to the finish. A further conclusion is that progress (i.e., the number of gaps remaining after screening n -fold genomes' worth of clones) is highly dependent on the percentage of DNA that must overlap between clones for the fingerprint to identify the overlap. Even with a detectable overlap as low as 20%, a genome the size of *E. coli* would be expected to require sevenfold coverage (about 2,000 lambda clones) to yield coverage with 10 gaps. With a detectable overlap of only 50%, 7-fold coverage yields 60 gaps, and 12-fold coverage is required to yield 10 gaps. "Gaps" in this context are not necessarily uncloned regions (gaps in the clone set) but rather lacks of continuity in the mapping of overlap relationships of the clones. They are informational gaps. These mapping gaps are sometimes called "apparent gaps" to distinguish them from "actual gaps," or cloning gaps.

Mathematically, the fingerprint strategy of isolating clone banks is completely analogous to random sequencing strategies, where overlap is detected by overlap in the sequence. A sequence is a very detailed physical description of DNA. Perhaps as few as 20 bases overlapping in an average sequence of 400 bases are sufficient to align two sequences. This is a 5% overlap, and it follows that a 10-fold random sequencing (some 120,000 random successful sequences) could theoretically yield the sequence of a genome the size of *E. coli* with only a handful of gaps.

These statistical predictions are based a number of simplifying assumptions that in actual practice do not hold. The library being screened is assumed to be a complete and random representation of the genome (i.e., there is no cloning bias). Deviation from randomness is the norm in libraries, and the effect on the model is to increase the coverage needed. The model also assumes that the structure of the DNA appears to be random and that the fingerprints occur and are distributed randomly. This is also not actually the case. Repeated regions, for example, are a problem. Despite these caveats, the mathematical model is a good starting point for evaluating expectations.

Fingerprints

Fingerprints used have either been a restriction map (an ordered set or sets of fragment lengths), an unordered set (or sets) of fragment lengths, or some combination. Fingerprints based on characteristics other than restriction site arrangements are also possible, and a method based on the pattern of hybridization to a defined set of oligonucleotide probes has been studied extensively (23, 37, 54, 65). A big advantage of this method is the potential for automation. Microchips of oligonucleotides could be made automatically, and hybridization results could be read automatically (35, 74). This type of fingerprinting scheme has been called a binary fingerprint, since each clone's fingerprint is the string of positive and negative scores with the set of oligonucleotide probes. Experimental parameters are the number of oligonucleotides to hybridize to the complete set of clones, the length of the oligonucleotide, and the specific oligonucleotides chosen (37). Experimentally, obtaining a fingerprint is not limited to cloned DNA. For example, gels of various restriction digests could be hybridized with the array of probes, and the fingerprint of each fragment could be obtained.

Many different schemes for using restriction enzymes to fingerprint clones have been used. Kohara et al. (48) used the maps of each of the clones for eight restriction enzymes. An average clone was 15 kb long and contained 25 mapped sites. Commonality of three to five map intervals was sufficient to define an overlap. With the Kohara et al. data set, an overlap of four intervals would be about 20% overlap.

We used a fingerprinting scheme based on the fragmentation pattern of six restriction digests (*EcoRI*, *BamHI*, *HindIII*, and all three double digests) for each clone. The fingerprint for each clone was information about all restriction fragments in each of the six digests: fragment size, the two enzymes that

define the fragment end points, and the binary information of whether or not there is a restriction site for the other enzymes between the fragment end points. We required at least two map intervals (three sites) in common before an overlap was assigned. This led to a few false overlaps. The average insert size was 16.3 kb, and the average number of fragments per clone was five; thus, the overlap detectable was about 40%. Of 2,000 analyzed random clones, 200 had three or fewer sites and thus could not overlap another clone in a useful way by this criterion.

A similar strategy called “landmark” (16) has been used to isolate encyclopedias for halobacteria (17, 91). In this scheme, the fingerprint (landmark) is a set of fragments resulting from complete digestion with a commonly occurring enzyme (e.g., 10 fragments per clone). Each fragment is distinguished by its size, the binary information of whether or not it is cut by each of 10 (or whatever number) rarely cutting landmark enzymes (e.g., one cut per clone), and, when possible, the sizes of the resulting subfragments.

Olson et al. (72), in isolating an overlapping set of clones for yeast cells, used the set of fragment lengths from one double digest (*EcoRI* + *HindIII*) for each clone. The overlap detectable was about 60%. Coulson et al. (22), for the nematode *Caenorhabditis elegans*, describe what has come to be generally meant when a fingerprint is referred to. Their fingerprint was the length of the small fragments obtained by labeling at rarely occurring sites (*HindIII*) and recutting at frequently occurring sites (*Sau3A*). An overlap of 40 to 50% was detectable.

Because of the need to fingerprint large numbers of clones, fingerprints that are experimentally easy to obtain have been chosen, and effort has gone into streamlining or automating fingerprinting techniques. The fingerprint schemes of both Olson et al. (72) and Coulson et al. (22) required only one digest and one lane of a gel per clone. Kohara’s fingerprint required eight partial digests and eight gel lanes per clone, blotting of each gel onto a membrane, and hybridization to a probe to visualize the map. Our fingerprint required larger amounts of pure DNA and six digests and gel lanes. The landmark of Charlebois et al. (16) required isolation of DNA for each clone, one complete digestion with the common enzyme, and up to 10 complete double digestions with the common enzyme plus another rarer (landmark) enzyme.

Experimentally, the methods also differ in the measurement error. Coulson et al. (22) measured small, end-labeled fragments on denaturing gels with an error of 1 to 4 bases depending on where in the gel the fragment migrated. We, Olson et al. (72), and Charlebois et al. (16) measured double-stranded fragments in the 0.5- to 20-kb range with a measurement error of 1 to 8%. Kohara et al. (48) measured fragments in the 8- to 30-kb range, and the fingerprint (map interval length) was actually the difference between these bands; thus, the error was quite large. Reducing the measurement error would increase the sensitivity by reducing the number of overlapping map intervals needed to define overlap.

A significant advantage of mapping-type fingerprinting schemes is that when the project is complete, one can obtain a complete map merely by merging the individual maps for all the clones. Thus, the encyclopedia is immediately useful for many purposes involving comparison with maps of other clones (83–85).

Hybridization Strategies

Strategies that use unique (single copy per genome) hybridization probes to define clone overlaps have several advantages, notably that the results of a probe-clone comparison are binary. False negatives, false positives, and ambiguities are due to experimental shortcomings rather than to the existence of the extremes in statistical distributions. Logistically, hybridization strategies are getting easier. Compact arrays of dot blots, automation in making dot blots, automated clone handling, databases for monitoring laboratory activities, new probe-making methods (e.g., primer extension, thermal cycling, STSs), and pooling schemes have greatly increased the ease of using hybridization to detect overlapping clones. Both top-down (walking) and bottom-up strategies have been employed (10, 102).

One can think of the unique probe hybridization methods as a spectrum between two extremes of probe size and probe-clone independence. At one extreme is the strategy of using the entirety of any clone as a probe against the entire set of clones (including itself). The positive and negative hybridization results

translate directly into positive and negative clone overlap relationships. The reciprocal experiment (switch the clone used as probe and the one used as target) gives the same result. The set of possible probes is identical to the set of clones. Two clones that both hybridize to the same probe do not necessarily overlap each other, since the probe has a significant length, and overlap with any part of it will lead to hybridization. They are, however, in the same “contig.”

At the other extreme is the use of short probes (e.g., genomic PCR products constituting an STS) distributed throughout the genome independently of the set of clones to be screened for overlap with each other. The short (relative to the length of the clones) probe identifies a subset of clones that contain the specific unique genomic site. Coincident clone hybridizations determine that the clones contain the same site and thus overlap. Nonhybridization, or differential hybridization, does not mean that the clones lack overlap but only that they do not overlap at the probe site. These strategies have been called “anchoring” (2, 31).

In most cases, the strategy under consideration is between these extremes. Often, the set of probes are not independent of the set of clones (for example, probes are sequences at each end of a clone); probes are not sites (e.g., they have significant length relative to the clones), or the probes are not independent of each other. Strategically, probes can be anonymous (e.g., a random library of short inserts in plasmids) or mapped (e.g., a set of previously studied clones or STSs), and they can be chosen (as the project progresses) to be derived from “unanchored” (i.e., nonoverlapping with each other) clones or from contig ends (walking) (2, 5, 31, 39, 70, 73, 102, 103).

With some information about the target genome and a hybridization-based strategy in mind, one can choose an n (number of randomly chosen clones) that is likely to be large enough to contain a complete detectable coverage of the genome. In general, n needs to be high enough to afford nearly 10-fold genome coverage. All n clones are processed and dot blotted (in duplicate to get enough membranes). The stepwise procedure is to choose a probe and hybridize it to the entire set of n clones. In each step, the data gathered will fill in some of the $n \times n$ clone overlap matrix and lead to the choice of the next probe. If the probes are clones themselves, then each step of hybridization fills in all rows in one column of the $n \times n$ matrix.

The number of probes needed to get enough data to order the clones can be greatly reduced by choosing the order of probes to be used on the basis of data gathered so far (103). Instead of using all clones to be probes, each new probe can be chosen to give new information. Each new probe is made from a clone not previously assigned to a contig; i.e., it has never yet overlapped any previously used probe. This procedure can be continued to about 0.7-fold probe coverage of the genome (sum of the lengths of clones from which probes are made), after which one expects to run out of clones that overlap nothing. (This is analogous to the “parking problem” or random sequential adsorption problem [38].) Using a variety of simplifying assumptions, one can predict statistically how many apparent gaps to expect during the course of the procedure (103).

Birkenbihl and Vielmetter (10) used a unique probe hybridization strategy to isolate an *E. coli* BHB2600 cosmid encyclopedia. Probes were nick translations of not-yet-overlapping cosmids from the clone sets. Cross-hybridization with a vector was suppressed with competing cold vector DNA. Those authors screened a total of 570 cosmid clones (5-fold coverage) with 65 cosmid clones (0.7-fold coverage). Near the end of the project, clones at the ends of contigs were used as probes to close five gaps, and *EcoRI* digests were used to verify the ordering into 12 contigs.

Strategies for Constructing Libraries

A major consideration in an encyclopedia project is choice of the vector system and cloning strategy to be used to produce a library. The cloning strategy should be such that the libraries from which random clones are chosen contain a complete set of randomly dispersed overlapping clones. The possible insert end points should occur frequently and be distributed randomly in the genome, so that no region is unclonable just because of a paucity of cloning sites. Use of more than one library is helpful. Selection against particular

regions should be minimized, and clones should be stable. Cloning multiple fragments should be avoided, as this leads to false adjacencies in the overlap data. These cloning artifacts have been called chimeras, rogues, or false contigs.

A minimal *E. coli* encyclopedia would require about 120 cosmid clones or 300 lambda phage clones. In general, cosmid and plasmid clones may accumulate deletions, because their rate of doubling depends inversely on DNA length and because the expression of cloned genes may be selected against because of deleterious metabolic imbalances. This could be especially true in the more useful high-copy-number cosmids. Phage lambda, on the other hand, destroys its host in a short time and thus need not coexist with it. Moreover, there is inherent selection against deletion in lambda clones because of the existence of a lower limit on the size of DNA that can be propagated. The lambda packaging limit can also be used effectively to reduce the number of false adjacencies created by cloning two fragments. Where possible, methods to verify that the libraries to be screened are representative should be employed before a mapping project is begun (10, 12).

Endgame: Strategies for Closing Gaps

At some point, it is reasonable to switch from a random approach to procedures directed at specifically covering gaps. In between the two processes, a number of procedures (such as correlation with the genetic map) can be used to organize and/or evaluate progress. Generally, gap closing involves screening libraries by hybridization to probes isolated from clones adjacent to the gap. If there are many gaps, the probe preparation may be as much effort as the initial fingerprinting or more. The appropriate point at which to switch might be estimated by balancing the predicted effort involved in each approach. In general, the amount of work for continuing the fingerprint is roughly proportional to the predicted number of clones that will have to be analyzed, and the effort per clone will be small. The effort to specifically close gaps is roughly proportional to the number of gaps, and the effort for each gap is large.

USES OF MAPS AND ENCYCLOPEDIAS

Before physical mapping methods were available, the study of bacterial chromosomes was limited to a few organisms for which genetic tools were available. The advent of physical mapping methods allows the detailed analysis of all microorganisms from which DNA can be obtained. The impact of these techniques is so sweeping that it is likely that genome analysis will begin with physical maps, and the use of genetic recombination to construct maps will be rare. Gene maps can be determined even in the absence of a genetic recombination tool by the hybridization of probes from related bacteria for various conserved genes (housekeeping enzymes, rRNA operons, etc.) to restriction fragments. Hybridization will assign the analogous gene to the restriction map.

Genome comparisons (size, gene order) are accomplished both by comparison of restriction maps (40) and by hybridization of fragments isolated from one mapped genome to digests of the other mapped genome. This kind of comparative analysis is becoming common as more and more bacterial genomes are mapped (20, 98) and is expected to yield a broad view of bacterial chromosome structure. For example, it is becoming clear that bacterial genomes are not always one circular molecule. Physical genome maps have shown that *Rhodobacter sphaeroides* (92), *Leptospira interrogans* (104), and *Brucella melitensis* (61) have two chromosomes. Linear chromosomes have been described for *Borrelia burgdorferi* (15, 69), and *Streptomyces lividans* (52). Genome size varies dramatically from the 600-kb genome of *Mycoplasma genitalium* (21) to the 8,000-kb genomes of *Streptomyces* spp. (52).

Specific gene mapping is also greatly facilitated by physical maps. Mapping of cloned genes (activities or phenotypes) is accomplished by physically mapping the clone and comparing its map to the high-resolution physical map of the genome (85). A cloned gene can also be mapped by hybridization of the clone to Southern blots of the mapped genome or to the complete clone bank (33, 66). Similarly, a gene

encoding a protein of interest can be readily located on the physical map by determining a small amount of amino acid sequence, designing an oligonucleotide probe (designed by reverse translation of the protein sequence), and probing genomic digests or a genomic encyclopedia.

An important application of gene encyclopedias is the mapping of mRNAs by hybridization of mRNA (cDNA) to the encyclopedia clones, thus identifying and mapping batteries of coregulated genes (19, 96). These methods complement the analysis by two-dimensional gels of whole bacterial protein content (64) and are a way to begin to study the multidimensional control networks from a global point of view.

Low-resolution maps facilitate classic genetic methods of isolating mutants and mapping mutations in related sets of genes. For example, gene identification and mapping by isolation of a phenotypically related set of mutants by transposon mutagenesis can be aided by using transposons with rare restriction sites for the mutagenesis. The mutations are physically mapped by restriction digestion and comparison of restriction pattern of mutant genome with the low-resolution restriction map of the parent genome.

In the same vein, genes identified by phenotype can be genetically mapped by rescue (or complementation) of mutant phenotypes by a clone in the clone bank or by transformation with an isolated DNA restriction fragment. Finally, clone bank members are a ready substrate for subcloning and sequencing.

LITERATURE CITED

1. **Amjad, M., J. M. Castro, H. Sandoval, J. J. Wu, M. Yang, D. J. Henner, and P. J. Piggot.** 1990. An *Sfi*I restriction map of the *Bacillus subtilis* 168 genome. *Gene* **101**:15–21.
2. **Arratia, R., E. S. Lander, S. Tavaré, and M. S. Waterman.** 1991. Genomic mapping by anchoring random clones: a mathematic analysis. *Genomics* **11**:806–827.
3. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. *Microbiol. Rev.* **54**:130–197.
4. **Bancroft, I., P. Wolk, and E. V. Oren.** 1989. Physical and genetic maps of the genome of the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC7120. *J. Bacteriol.* **171**:5940–5948.
5. **Barillot, E., J. Dausset, and D. Cohen.** 1991. Theoretical analysis of a physical mapping strategy using random single-copy landmarks. *Proc. Natl. Acad. Sci. USA* **88**: 3917–3921.
6. **Bautsch, W.** 1988. Rapid physical mapping of the *Mycoplasma mobile* by two-dimensional field inversion gel electrophoresis techniques. *Nucleic Acids Res.* **16**:11461–11467.
7. **Bautsch, W.** 1993. A *Nhe*I macrorestriction map of the *Neisseria meningitidis* B1940 genome. *FEMS Microbiol. Lett.* **107**:191–198.
8. **Berlyn, M. B., and S. Letovsky.** 1992. Genome-related datasets within the *E. coli* Genetic Stock Center database. *Nucleic Acids Res.* **20**:6143–6151.
9. **Bingham, N. W. K., R. Khawaja, H. Louie, E. Hani, K. Neote, and V. L. Chan.** 1992. Physical map of *Campylobacter jejuni* TGH9011 and localization of 10 genetic markers by use of pulsed-field gel electrophoresis. *J. Bacteriol.* **174**:3494–3498.
10. **Birkenbihl, R. P., and W. Vielmetter.** 1989. Cosmid-derived map of *E. coli* strain BHB2600 in comparison to the map of strain W3110. *Nucleic Acids Res.* **17**:5057–5069.
11. **Borges, K. M., and P. L. Bergquist.** 1993. Genomic restriction map of the extremely thermophilic bacterium *Thermus thermophilus* HB8. *J. Bacteriol.* **175**:103–110.
12. **Brody, H., J. Griffith, A. J. Cuticchia, J. Arnold, and W. E. Timberlake.** 1991. Chromosome-specific recombinant DNA libraries from the fungus *Aspergillus nidulans*. *Nucleic Acids Res.* **19**:3105–3109.
13. **Carlson, C. R., A. Grønstad, and A.-B. Kolstø.** 1992. Physical maps of the genomes of three *Bacillus cereus* strains. *J. Bacteriol.* **174**:3750–3756.
14. **Carlson, C. R., and A.-B. Kolstø.** 1993. A complete physical map of a *Bacillus thuringiensis* chromosome. *J. Bacteriol.* **174**:1053–1060.
15. **Casjens, S., and W. M. Huang.** 1993. Linear chromosomal physical and genetic map of *Borrelia*

- burgdorferi*, the Lyme disease agent. *Mol. Microbiol.* **8**:967–980.
16. **Charlebois, R. L., J. D. Hofman, L. C. Schalkwyk, W. L. Lam, and W. F. Doolittle.** 1989. Genome mapping in halobacteria. *Can. J. Microbiol.* **35**:21–29.
 17. **Charlebois, R. L., L. C. Schalkwyk, J. D. Hofman, and W. F. Doolittle.** 1991. Detailed physical map and set of overlapping clones covering the genome of the archaebacterium *Haloferax volcanii* DS2. *J. Mol. Biol.* **222**:509–524.
 18. **Chen, X., and W. R. Widger.** 1993. Physical genome map of the unicellular cyanobacterium *Synechococcus* sp. strain PCC 7002. *J. Bacteriol.* **175**:5106–5116.
 19. **Chuang, S.-E., D. L. Daniels, and F. R. Blattner.** 1993. Global regulation of gene expression in *Escherichia coli*. *J. Bacteriol.* **175**:2026–2036.
 20. **Cole, S. T., and I. Saint Girons.** 1994. Bacterial genomics. *FEMS Microbiol. Rev.* **14**:139–160.
 21. **Colman, S. D., P. C. Hu, W. Litaker, and K. F. Bott.** 1990. A physical map of the *Mycoplasma genitalium* genome. *Mol. Microbiol.* **4**:683–687.
 22. **Coulson, A., J. Sulston, S. Brenner, and J. Karn.** 1986. Toward a physical map of the genome of the nematode, *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **83**:7821–7825.
 23. **Cuticchia, A. J., J. Arnold, H. Brody and W. E. Timberlake.** 1992. CMAP: contig mapping and analysis package, a relational database for chromosome reconstruction. *Comput. Appl. Biosci.* **8**:467–474.
 24. **Daniels, D. L.** 1989. AvrII maps of *E. coli* strains. *Nucleic Acids Res.* **19**:589–597.
 25. **Daniels, D. L.** 1990. Constructing encyclopedias of genomes, p. 43–51. In K. Drlica and M. Riley (ed.), *The Bacterial Chromosome*. American Society for Microbiology, Washington, D.C.
 26. **Daniels, D. L., and F. R. Blattner.** 1987. Mapping using gene encyclopaedias. *Nature* (London) **325**:831–832.
 27. **Dingwall, A., and L. Shapiro.** 1989. Rate, origin, and bidirectionality of *Caulobacter* chromosome replication as determined by pulsed-field gel electrophoresis. *Proc. Natl. Acad. Sci. USA* **86**:119–123.
 28. **Eiglmeier, K., N. Honoré, S. A. Woods, B. Caudron, and S. T. Cole.** 1993. Use of an ordered cosmid library to deduce the genomic organization of *Mycobacterium leprae*. *Mol. Microbiol.* **7**:197–206.
 29. **Ely, B., T. W. Ely, C. J. Gerardot, and A. Dingwall.** 1990. Circularity of the *Caulobacter crescentus* chromosome determined by pulsed-field gel electrophoresis. *J. Bacteriol.* **172**:1262–1266.
 30. **Ely, B., and C. J. Gerardot.** 1988. Use of pulsed-field gradient gel electrophoresis to construct a physical map of the *Caulobacter crescentus* genome. *Gene* **68**:323–333.
 31. **Ewens, W. J., C. J. Bell, P. J. Donnelly, P. Dunn, E. Matallana, and J. R. Ecker.** 1991. Genome mapping with anchored clones: theoretical aspects. *Genomics* **11**:799–805.
 32. **Fan, J. B., Y. Chikashige, C. L. Smith, O. Niwa, M. Yanagida, and C. R. Cantor.** 1988. Construction of a *NotI* restriction map of the fission yeast *Schizosaccharomyces pombe* genome. *Nucleic Acids Res.* **17**:2801–2818.
 33. **Farinha, M. A., S. L. Ronald, A. M. Kropinski, and W. Paranchych.** 1993. Localization of the virulence-associated genes *pilA*, *pilR*, *rpoN*, *fliA*, *fliC*, *ent*, and *fbp* on the physical map of *Pseudomonas aeruginosa* PAO1 by pulsed-field electrophoresis. *Infect. Immun.* **61**:1571–1575.
 34. **Fiandt, M., Z. Hradecna, H. A. Lozeron, and W. Szybalski.** 1971. Electron micrographic mapping of deletions, insertions, inversions, and homologies in the DNAs of coliphages lambda and phi80, p. 329–354. In A. D. Hershey (ed.), *The Bacteriophage Lambda*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 35. **Fodor, S. P. A., J. L. Read, M. C. Pirrung, L. Stryer, A. T. Lu, and D. Solas.** 1991. Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**:767–773.
 36. **Fonstein, M., and R. Haselkorn.** 1993. Chromosomal structure of *Rhodobacter capsulatus* strain SB1003: cosmid encyclopedia and high-resolution physical and genetic map. *Proc. Natl. Acad. Sci. USA* **90**:2522–2526.

37. **Fu, Y. X., W. E. Timberlake, and J. Arnold.** 1992. On the design of genome mapping experiments using short synthetic oligonucleotides. *Biometrics* **48**:337–359.
38. **Gonzalez, J. J., P. C. Hemmer, and J. S. Hoye.** 1974. Cooperative effects in random sequential polymer reactions. *Chem. Phys.* **3**:228–238.
39. **Grigoriev, A. V.** 1993. Theoretical predictions and experimental observations of genomic mapping by anchoring random clones. *Genomics* **15**:311–316.
40. **Grothues, D., and B. Tümmler.** 1991. New approaches in genome analysis by pulsed-field gel electrophoresis: application to the analysis of *Pseudomonas* species. *Mol. Microbiol.* **5**:2763–2776.
41. **Hanish, J., and M. McClelland.** 1990. Methylase-limited partial *NotI* cleavage for physical mapping of genome DNA. *Nucleic Acids Res.* **18**:3287–3291.
42. **Heath, J. D., J. D. Perkins, B. Sharma, and G. M. Weinstock.** 1992. *NotI* genomic cleavage map of *Escherichia coli* K-12 strain MG1655. *J. Bacteriol.* **174**:558–567.
43. **Itaya, M., and T. Tanaka.** 1991. Complete physical map of the *Bacillus subtilis* 168 chromosome constructed by a gene-directed mutagenesis method. *J. Mol. Biol.* **220**:631–648.
44. **Kauc, L., M. Mitchell, and S. H. Goodgal.** 1989. Size and physical map of the chromosome of *Haemophilus influenzae*. *J. Bacteriol.* **171**:2474–2479.
45. **Kieser, H. M., T. Kieser, and D. A. Hopwood.** 1992. A combined genetic and physical map of the *Streptomyces coelicolor* A3(2) chromosome. *J. Bacteriol.* **174**:5496–5507.
46. **Knott, V., D. J. Rees, Z. Cheng, and G. G. Brownlee.** 1988. Randomly picked cosmid clones overlap the *pyrB* and *oriC* gap in the physical map of the *E. coli* chromosome *Nucleic Acids Res.* **16**:2601–2612.
47. **Kohara, Y.** 1990. Correlation between the physical and genetic maps of the *Escherichia coli* K-12 chromosome, p. 29–42. In K. Drlica and M. Riley (ed.), *The Bacterial Chromosome*. American Society for Microbiology, Washington, D.C.
48. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the W3110 *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50**:495–508.
49. **Kolstø, A.-B., A. Grønstad, and H. Oppegaard.** 1990. Physical map of the *Bacillus cereus* chromosome. *J. Bacteriol.* **172**:3821–3825.
50. **Kundig, C., H. Hennecke, and M. Göttfert.** 1993. Correlated physical and genetic map of the *Bradyrhizobium japonicum* 110 genome. *J. Bacteriol.* **175**:613–622.
51. **Lander, E. S., and M. S. Waterman.** 1988. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**:231–239.
52. **LeBlond, P., M. Redenbach, and J. Cullum.** 1993. Physical map of the *Streptomyces lividans* 66 genome and comparison with that of the related strain *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **175**:3422–3429.
53. **Lee, J. J., H. O. Smith, and R. J. Redfield.** 1989. Organization of the *Haemophilus influenzae* Rd genome. *J. Bacteriol.* **171**:3016–3024.
54. **Lehrach, H., R. Drmanac, J. Hoheisel, Z. Larin, G. Lennon, A. P. Monaco, D. Nizetic, and A. Poutska.** 1990. Hybridization fingerprinting in genome mapping and sequencing. In K. E. Davies and S. M. Tilghman (ed.), *Genome Analysis*, vol. 1. *Genetic and Physical Mapping*. Cold Spring Harbor Laboratory Press, Plainview, N.Y.
55. **Liu, S.-L., A. Hessel, H.-Y. M. Cheng, and K. E. Sanderson.** 1994. The *XbaI-BlnI-CeuI* genomic cleavage map of *Salmonella paratyphi* B. *J. Bacteriol.* **176**:1014–1024.
56. **Liu, S.-L., A. Hessel, and K. E. Sanderson.** 1993. The *XbaI-BlnI-CeuI* genomic cleavage map of *Salmonella typhimurium* LT2 determined by double digestion, end labeling, and pulsed-field gel electrophoresis. *J. Bacteriol.* **175**:4104–4120.
57. **Liu, S.-L., and K. E. Sanderson.** 1992. A physical map of the *Salmonella typhimurium* LT2 genome made by using *XbaI* analysis. *J. Bacteriol.* **174**:1662–1672.
58. **López-García, P., J. P. Abad, C. Smith, and R. Amils.** 1992. Genomic organization of the

- halophilic archaeon *Haloferax mediterranei*: physical map of the chromosome. *Nucleic Acids Res.* **20**:2459–2464.
59. **McGuigan, T. L., K. J. Livak, and S. Brenner.** 1993. DNA fingerprinting by sampled sequencing. *Methods Enzymol.* **218**:241–258.
60. **Medigue, C., A. Viarai, A. Henaut, and A. Danchin.** 1993. Colibri: a functional data base for the *Escherichia coli* genome. *Microbiol. Rev.* **57**:623–654.
61. **Michaux, S., J. Paillisson, M.-J. Carles-Nurit, G. Bourg, A. Allardet-Servent, and M. Ramuz.** 1993. Presence of two independent chromosomes in the *Brucella melitensis* 16M genome. *J. Bacteriol.* **175**:701–705.
62. **Michel, E., and P. Cossart.** 1992. Physical map of the *Listeria monocytogenes* chromosome. *J. Bacteriol.* **174**:7098–7103.
63. **Murray, B. E., K. V. Singh, R. P. Ross, J. D. Heath, G. M. Dunny, and G. M. Weinstock.** 1993. Generation of restriction map of *Enterococcus faecalis* OG1 and investigation of growth requirements and regions encoding biosynthetic function. *J. Bacteriol.* **175**:5216–5223.
64. **Neidhardt, F. C.** 1987. Multigene systems and regulons, p. 1313–1317. In F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, vol. 2. American Society for Microbiology, Washington, D.C.
65. **Newberg, L. A.** 1994. Finding a most likely clone ordering from oligonucleotide hybridization data. *Genomics* **21**:602–611.
66. **Noda, A., J. B. Courtright, P. F. Denor, G. Webb, Y. Kohara, and A. Ishihama.** 1991. Rapid identification of specific genes in *E. coli* by hybridization to membranes containing the ordered set of phage clones. *BioTechniques* **10**:474–476.
67. **Noll, K. M.** 1989. Chromosome map of the thermophilic archaeobacterium *Thermococcus celer*. *J. Bacteriol.* **171**:6720–6725.
68. **Okada, N., C. Sasakawa, T. Tobe, K. A. Talukder, K. Komatsu, and M. Yoshikawa.** 1991. Construction of a physical map of the chromosome of *Shigella flexneri* 2a and the direct assignment of nine virulence-associated loci identified by Tn5 insertions. *Mol. Microbiol.* **5**:2171–2180.
69. **Old, I. G., J. Mac Dougall, I. Saint Girons, and B. E. Davidson.** 1992. Mapping of genes on the linear chromosome of the bacterium *Borrelia burgdorferi*: possible locations for its origin of replication. *FEMS Microbiol. Lett.* **99**:245–250.
70. **Olson, M. V., and P. Green.** 1993. Criterion for the completeness of large-scale physical maps of DNA. *Cold Spring Harbor Symp. Quant. Biol.* **58**:349–355.
71. **Olson, M., L. Hood, C. Cantor, and D. Botstein.** 1989. A common language for physical mapping of the human genome. *Science* **245**:1434.
72. **Olson, M. V., J. E. Dutchik, M. Y. Graham, G. M. Brodeur, C. Helms, M. Frank, M. MacCollin, R. Scheinman, and T. Frank.** 1986. Random-clone strategy for genomic restriction mapping in yeast. *Proc. Natl. Acad. Sci. USA* **83**:7826–7830.
73. **Palazzolo, M. J., S. A. Sawyer, C. H. Martin, D. A. Smoller, and D. L. Hartl.** 1991. Optimized strategies for sequence-tagged-site selection in genome mapping. *Proc. Natl. Acad. Sci. USA* **88**:8034–8038.
74. **Pease, A. C., D. Solas, E. J. Sullivan, M. T. Cronin, C. P. Holmes, and S. P. A. Fodor.** 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl. Acad. Sci. USA* **91**:5022–5026.
75. **Perkins, J. D., J. D. Heath, B. R. Sharma, and G. M. Weinstock.** 1992. *Sfi*I genomic cleavage map of *Escherichia coli* K-12 strain MG1655. *Nucleic Acids Res.* **20**:1129–1137.
76. **Perkins, J. D., J. D. Heath, B. R. Sharma, and G. M. Weinstock.** 1993. *Xba*I and *Bln*I genomic cleavage maps of *Escherichia coli* K-12 strain MG1655 and comparative analysis of other strains. *J. Mol. Biol.* **232**:419–445.

77. **Poustka, A.** 1993. Construction and use of chromosome jumping libraries. *Methods Enzymol.* **217**:358–378.
78. **Pyle, L. E., and L. R. Finch.** 1988. Preparation and FIGE separation of infrequent restriction fragments from *Mycoplasma mycoides* DNA. *Nucleic Acids Res.* **16**:2263–2268.
79. **Pyle, L. E., T. Taylor, and L. R. Finch.** 1990. Genomic maps of some strains within the *Mycoplasma mycoides* cluster. *J. Bacteriol.* **172**:7265–7268.
80. **Rackwitz, H. R., A. M. Zehetner, A. M. Frischauf, and H. Lehrach.** 1984. Rapid restriction mapping of DNA cloned in lambda phage vectors. *Gene* **30**:195–200.
81. **Ratnaningsih, E., S. Dharmsthiti, V. Krishnapillai, A. Morgan, M. Sinclair, and B. W. Holloway.** 1990. A combined physical and genetic map of *Pseudomonas aeruginosa* PAO. *J. Gen. Microbiol.* **136**:2351–2357.
82. **Romling, U., D. Grothues, W. Bautsch, and B. Tummeler.** 1989. A physical genome map of *Pseudomonas aeruginosa* PAO. *EMBO J.* **8**:4081–4089.
83. **Rudd, K. E.** 1992. Alignment of *E. coli* DNA sequences to a revised, integrated genomic restriction map, p. 2.3–2.43. In J. Miller (ed.), *A Short Course in Bacterial Genetics: a Laboratory Manual and Handbook for Escherichia coli and Related Bacteria*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.
84. **Rudd, K. E., W. Miller, J. Ostell, and D. A. Benson.** 1990. Alignment of *Escherichia coli* K12 DNA sequences to a genomic restriction map. *Nucleic Acids Res.* **18**:313–321.
85. **Rudd, K. E., W. Miller, C. Werner, J. Ostell, C. Tolstoshev, and S. G. Satterfield.** 1991. Mapping sequenced *E. coli* genes by computer: software, strategies and examples. *Nucleic Acids Res.* **19**:637–647.
86. **Schroeder, J. L., and F. R. Blattner.** 1978. Least-squares method for restriction mapping. *Gene* **4**:169–174.
87. **Schwartz, D. C., and C. R. Cantor.** 1984. Separation of yeast chromosome-sized DNAs by pulsed-field gel electrophoresis. *Cell* **37**:67–75.
88. **Smith, C. L., G. Condemine, S. Y. Chang, E. McGary, and S. Chang.** 1987. A physical map of the *Escherichia coli* K12 genome. *Science* **236**:1448–1453.
89. **Smith, C. L., P. E. Warburton, A. Gaal, and C. R. Cantor.** 1986. Analysis of genome organization and rearrangements by pulsed field gradient gel electrophoresis, p. 45–70. In J. K. Setlow and A. Hollaender (ed.), *Genetic Engineering* 8. Plenum, New York.
90. **Smith, H. O., and M. L. Birnstein.** 1976. A simple method for DNA restriction site mapping. *Nucleic Acids Res.* **3**:2387–2398.
91. **St. Jean, A., B. A. Trieselmann, and R. L. Charlebois.** 1994. Physical map and set of overlapping cosmid clones representing the genome of the archaeon *Halobacterium* sp. GRB. *Nucleic Acids Res.* **22**:1476–1483.
92. **Suwanto, A., and S. Kaplin.** 1989. Physical and genetic mapping of the *Rhodobacter sphaeroides* 2.4.1 genome: presence of two unique circular chromosomes. *J. Bacteriol.* **171**:5850–5859.
93. **Szybalski, E. H., and W. Szybalski.** 1982. A physical map of the *Escherichia coli* bio operon. *Gene* **19**:93–104.
94. **Tabata, S., A. Higashitani, 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.
95. **Taylor, D. E., M. Eaton, N. Chang, and S. M. Salama.** 1992. Construction of a *Helicobacter pylori* genome map and demonstration of diversity at the genome level. *J. Bacteriol.* **174**:6800–6806.
96. **Trieselmann, B. A., and R. L. Charlebois.** 1992. Transcriptionally active regions in the genome of the archaeobacterium *Haloferax volcanii*. *J. Bacteriol.* **174**: 30–34.
97. **Ventra, L., and A. S. Weiss.** 1989. Transposon-mediated restriction mapping of the *Bacillus subtilis* chromosome. *Gene* **78**:29–36.
98. **Weinstock, G. M.** 1994. Bacterial genomes: mapping and stability. *ASM News* **60**:73–78.

99. **Wenzel, R., and R. Herrmann.** 1988. Physical mapping of the *Mycoplasma pneumoniae* genome. *Nucleic Acids Res.* **16**:8323–8336.
100. **Wenzel, R., and R. Herrmann.** 1989. Cloning of the complete *Mycoplasma pneumoniae* genome. *Nucleic Acids Res.* **17**:7029–7043.
101. **Wong, K. K., and M. McClelland.** 1992. A *BlnI* restriction map of the *Salmonella typhimurium* LT2 genome. *J. Bacteriol.* **174**:1656–1661.
102. **Yoshida, K., M. P. Strathmann, C. A. Mayeda, C. H. Martin, and M. J. Palazzolo.** 1993. A simple and efficient method for constructing high resolution physical maps. *Nucleic Acids Res.* **21**:3553–3562.
103. **Zhang, M. Q., and T. G. Marr.** 1993. Genome mapping by nonrandom anchoring: a discrete theoretical analysis. *Proc. Natl. Acad. Sci. USA* **90**:600–604.
104. **Zuerner, R., J.-L. Herrmann, and I. Saint Girons.** 1993. **Comparison of genetic maps for two *Leptospira interrogans* serovars provides evidence for two chromosomes and intraspecies heterogeneity.** *J. Bacteriol.* **175**:5445–5451.
105. **Zuerner, R. L., and T. B. Stanton.** 1994. Physical and genetic map of the *Serpulina hyodysenteriae* B78T chromosome. *J. Bacteriol.* **176**:1087–1092.