Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 5030–5034, July 1989
Genetics

Physical mapping of complex genomes by cosmid multiplex analysis

(genomic cloning/chromosome 11/genomic data base/human genome/DNA fingerprinting)

G. A. EVANS* AND K. A. LEWIS

Molecular Genetics Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037

Communicated by Renato Dulbecco, March 27, 1989 (received for review January 20, 1989)

ABSTRACT **A rapid and powerful approach for linking individual clones of a cosmid library and the assembly of a large physical map is presented, which depends on the simultaneous analysis of many cosmid clones for overlapping regions. This method uses cosmid vectors that contain endogenous bacteriophage T3 and T7 promoters to allow for the identification of overlapping clones through the synthesis of end-specific RNA probes. A genomic library is constructed and organized as an ordered matrix such that each clone is assigned an identifying coordinate. DNA from mixtures of cosmid clones is pooled such that each pool contains only one common member with any other pool, RNA probes are prepared from mixtures of cosmid clones, and groups of clones overlapping with the constituents of the mixtures are determined by hybridization. Pooled probes are most simply prepared by grouping clones according to the rows and columns of the library matrix. The pairwise comparison of data generated by the hybridization of mixed probes can be decoded by using simple algorithms that predict the order and linkage of all clones in the collection and organize them into predicted contigs. To demonstrate the feasibility of multiplexed analysis of cosmids, a genomic library was prepared from a mouse–human somatic cell hybrid that contains a portion of the long arm of human chromosome 11. Preparation, arrangement on a matrix, and analysis of pooled cosmid clones from this collection resulted in the detection of 1099 linked pairs of cosmids, which could be assembled into 315 contigs. Thus, with a minimal amount of effort, a substantial portion of this genomic region has been linked in multiple overlapping contigs. This method may have practical applications in the large-scale mapping and sequencing of mammalian genomes.**

The analysis of large genomes will require the application of both "top-down" and "bottom-up" mapping strategies. The former strategy depends on the separation on pulsed-field gels of large DNA fragments generated by using rare restriction endonucleases for physical linkage of DNA markers and the construction of long-range maps (1–4). The latter strategy depends on identifying overlapping sequences in a large number of randomly selected bacteriophage or cosmid clones by unique restriction enzyme "fingerprinting" (4, 5) and their assembly into overlapping sets of clones referred to as contigs (6). A similar strategy is potentially feasible using megabase-size fragments cloned as yeast artificial chromosomes (3). In the past few years, bottom-up mapping strategies have been successfully applied to generate complete or partial genomic maps of *Saccharomyces cerevisiae* (4), *Caenorhabditis elegans* (5, 7), and *Escherichia coli* (8). Poustka *et al.* (9) proposed a different strategy for the ordering of cosmid or phage clones organized at high density on a matrix by using identifying sequences detected with short oligonucleotide probes.

In this paper, we describe an alternate strategy for bottom-up mapping that is applicable to the analysis of mammalian chromosomes and allows for the simultaneous determination of overlaps between cosmid clones analyzed in pools. Rather than depending on fingerprinting procedures for detection of overlapping clones, we constructed cosmid libraries by using vectors containing T3 or T7 bacteriophage promoters flanking the cloned genomic DNA (10, 11) such that overlapping sequences could be detected by hybridization. To test the feasibility of this strategy, we analyzed 960 clones isolated from human chromosome 11 for overlapping regions by preparing 68 mixed RNA probes, which were used for hybridization to replicas of this filter. This simple procedure resulted in the ordering of cosmid clones spanning 11q13–11qter and the identification of contigs containing many of the genes mapping to this chromosome. This method is potentially applicable to cloning, ordering clones, and physical mapping of other complex genomes as well.

MATERIALS AND METHODS

Cell Lines. TG 5D1-1 is a Friend cell line derived from somatic cell hybrid 5D1 [which carries an intact human X chromosome and chromosome 11 (12)] and selected with 6-thioguanine for the loss of the entire X chromosome and most of chromosome 11. TG 5D1-1 contains the distal portion of chromosome 11 as the only human material in a mouse genomic background (13), representing about 0.9% of the human genome.

Cosmid Vectors and Libraries. Genomic libraries were constructed in cosmid vector sCos-1 (11), which contains duplicated *cos* sites for high efficiency microcloning, T3 and T7 bacteriophage promoters flanking the unique *Bam*HI cloning site, two *Not* I sites for the excision of genomic inserts, a selectable gene (SV2-neo^r) for mammalian gene transfer, and a ColE1 origin of replication (see Fig. 1). Detailed restriction maps of the cosmid insert in this vector may be rapidly determined by an end-labeling mapping procedure using T3- or T7-specific oligonucleotides (11, 14, 15). The genomic cosmid library used in this study consisted of 1.5×10^7 independent clones and was constructed by using genomic DNA digested to an average size of 100–120 kilobases, dephosphorylated with calf intestinal phosphatase, and packaged with Gigapak Gold (Stratagene) *in vitro* packaging lysate (11). Only nonamplified libraries were used, and cosmid clones were archived in 96-well microtiter plates stored at −70°C in LB media with 15% (vol/vol) glycerol and kanamycin sulfate at 25 μg/ml.

Library Screening. Cosmid libraries were plated on 576-cm^2 LB agar trays at a density of 10 clones per cm^2, replica filters were prepared, and filters were hybridized with human placenta DNA labeled with [^{32}P]dCTP to a specific activity of 10^8 cpm/μg. Under these hybridization conditions, no back-

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

*To whom reprint requests should be addressed at: Molecular Genetics Laboratory, The Salk Institute, P.O. Box 85800, San Diego, CA 92138.

ground hybridization was detected against cosmids carrying mouse genomic DNA. Cosmids containing human genomic DNA inserts were picked with toothpicks, rescreened by hybridization to ^{32}P-labeled human DNA, and transferred to 96-well archive microtiter plates. Replica transfer of clones in 96-well microtiter plates and transfer from archived plates to screening filters was carried out with an aluminum "hedgehog" (5) or a simple laboratory robot (Beckman Biomek 1000). For multiplex analysis, archived cosmids were inoculated on the surface of a nitrocellulose- or nylon-based filter in a matrix or grid pattern using a 36 × 36 matrix based on the size and spacing of the 96-well archive plates. The clones were allowed to grow on the surface of the filter at 37°C for 12–15 hr, and bacterial DNA was fixed to the filter by using a standard colony lysis procedure (16).

RNA Probe Synthesis and Hybridization Reactions. Cosmids were transferred from archives to fresh 96-well plates containing liquid LB media with kanamycin sulfate at 25 μg/ml and incubated at 37°C in a humidified atmosphere for 6–10 hr. Supernatants from individual wells were pooled, and DNA was prepared by using a cosmid minilysate procedure (14). RNA probes were synthesized as previously described using bacteriophage T3 or T7 polymerase (Stratagene), and [^{32}P]UTP and polymerase reactions were terminated by extraction with phenol and chloroform. The RNA probe was prehybridized with a blocking mixture (a mixture of sonicated human placenta DNA and cloned human repetitive sequences at a concentration of 1 mg/ml) as described (11) and then hybridized to a replica of the matrix filter for 12–18 hr under previously described conditions (11). Filters were washed in 0.1× SSC (0.015 M NaCl/0.0015 M sodium citrate, pH 7.6)/0.1% SDS at 65°C and exposed to x-ray film for 2–8 hr. Restriction enzyme analysis of isolated cosmids were carried out by using labeled oligonucleotides recognizing the T3 or T7 promoter sequences as described (11, 15).

Data Analysis. The grid coordinates of hybridizing cosmid clones with each pool of probes were entered into a computer file and analyzed by using computer programs written by G.A.E. in Turbo Pascal (Borland International) running on Apple Macintosh II or Macintosh SE computers. These programs compared data sets from hybridization reactions using different probe pools, identified those clones that were detected by more than one probe mixture, produced a list of linked clones, and assembled the list of overlapping clones into potential contigs using a simple tree-building algorithm. In some cases, orientation and overlap of individual cosmid clones in a contig were confirmed by detailed restriction mapping and hybridization analysis of the individual cosmid clones.

RESULTS

Cosmid Multiplex Mapping. We reasoned that significant improvements in the speed and efficiency of bottom-up genomic mapping using cosmid clones could be achieved if (*i*) restricted regions of large mammalian genomes could be isolated in a sublibrary organized as a matrix on a solid filter support, (*ii*) hybridization of end-specific probes could be used for the detection of overlaps rather than fingerprinting followed by pattern recognition, and (*iii*) multiple clones could be analyzed simultaneously for the detection of all possible overlaps in the collection. A theoretical analysis of fingerprinting techniques has suggested that the efficiency of the analysis is strongly dependent on the criteria used to declare overlaps between clones (17), and previously described methods (4, 5, 8) require 25-50% overlap between contiguous clones to allow detection. The use of end-specific RNA probes, synthesized from cosmid vectors containing bacteriophage T3 and T7 promoters (Fig. 1), allow for the unambiguous detection of overlapping regions as small as 0.5% (10) and provide strategies for the simultaneous analysis of pools of clones.

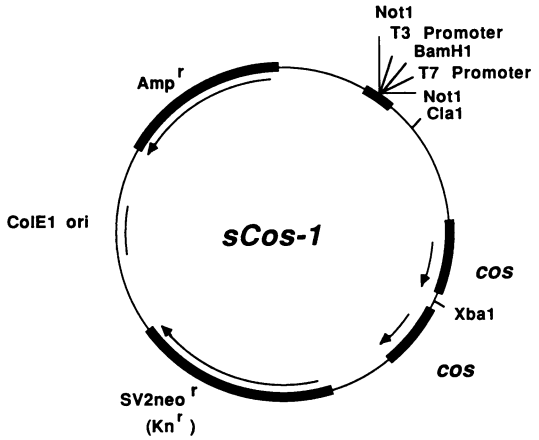

FIG. 1. A vector for cosmid multiplex analysis. The vector sCos-1 contains bacteriophage T3 and T7 promoters flanking a unique *Bam*HI cloning site, *Not* I sites for expedited restriction mapping and excision of the insert DNA, duplicated *cos* sites for high efficiency microcloning, a dominant selection for transfection into mammalian cells (SV2neo^r), ampicillin (Amp^r) and kanamycin (Kn^r) resistance genes, and ColE1 origin of replication (11).

Overlapping contiguous cosmid clones arranged on an organized matrix may be detected by the synthesis of an end-specific RNA probe and hybridization of the probe to a replica of the matrix filter. In addition, the use of hybridization to detect overlapping clones allows a multiplex strategy where the RNA probes are prepared from predetermined pools of cosmid templates. The templates are pooled such that each two pools contain only one cosmid in common; thus, comparison of the results of hybridization of two different probes to a matrix filter will ensure that clones detected by both pools represent that hybridizing to the common clone. A simple way to prepare these mixed probes is to pool all of the cosmid clones corresponding to a row of the two-dimensional matrix, prepare end-specific RNA probes, and carry out a hybridization reaction to a replica of the organized matrix. Each probe will hybridize to its own template resulting in the detection of a complete row and, in addition, a collection representing all of the clones overlapping with templates (Fig. 2). A probe prepared from the pool of cosmids representing a column detects a similar pattern. Those hybridizing clones that appear in both data sets but are not in a template row or column result from hybridization of the common template and indicate physical overlap with the clone that is located at the intersection of the template row and column of the matrix. Thus, if probes are prepared from pools of all of the rows and columns, a large number of overlapping clones in the collection can be rapidly detected. Thus, N clones organized in a two-dimensional array could be analyzed using $2N^{1/2}$ reactions.

Analysis of Human Chromosome 11q. To test the feasibility of this strategy for mapping portions of the human genome, we used cosmid vector sCos-1 (Fig. 1) to prepare a genomic library from a somatic cell hybrid containing as its only human material DNA from the distal long arm of human chromosome 11, including 11q13–11qter, in a mouse genomic background (13). The proportion of human clones in this genomic library was 0.9%, indicating that the hybrid cell line carried about 27 megabases of the human genome, consistent with previous cytogenetic and molecular characterization (13). By screening with labeled human placenta DNA, 960 cosmids containing exclusively human genomic DNA were selected, archived in 96-well microtiter plates, and arranged on a nitrocellulose filter according to the rows and columns

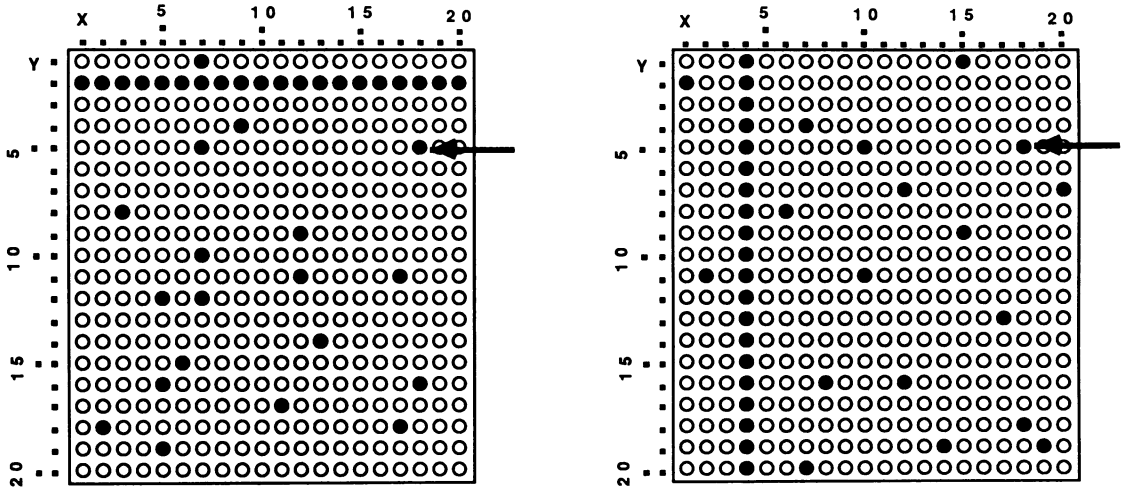

FIG. 2. Strategy for analysis of physical linkage by using groups of cosmids. Cosmid clones are inoculated on the surface of a nitrocellulose or nylon filter from 96-well archive plates stored at −70°C. Each clone on the "grid" is assigned a unique identifying *y*- and *x*-axis coordinate. Individual clones in the collection contain the innate capacity of generating probes specific for the extreme ends of the genomic DNA insert and detecting overlapping clones on the filter. To enable analysis of multiple clones simultaneously, cosmids are pooled according to the rows and columns of the matrix, DNA is prepared, and a mixed RNA probe is synthesized. When hybridized to the matrix filter, the probe detects a pattern of spots consisting of all of the template clones and the collection of clones overlapping with one end of each of the template clones. A similar procedure is carried out by using cosmids pooled according to columns of the matrix. When the two data sets are compared, hybridizing clones identified by both of the mixed probes may be overlapping with the template clone common to both sets: that clone is located at the intersection of the row and column. This procedure may then be repeated with other combinations of pooled templates and either T7 or T3 polymerase-derived probes. The arrows denote the location of a clone that overlaps with the "T7 end" of the clone at coordinates $y = 2$, $x = 4$.

of a 36 × 36 matrix (though not completing the 1296 member matrix). This collection is about 1.0 to 1.5 times redundant for this portion of chromosome 11 and is adequate for testing the multiplex strategy, recognizing that a far greater representation would be required to generate a complete cosmid set without gaps. Cosmids containing genes previously mapped to chromosome 11q were detected with available probes, and the genes *THY1* (18), *T3D*, *T3E* (15), *ETS1* (19), *PBG* (20), *PGR* (21), *SRPR* (22), and *APOA1* (23) were assigned unique *y* and *x* coordinates.

Multiplex analysis was carried out as follows: Cosmids were pooled according to 32 rows and 36 columns, RNA probes were prepared using T7 polymerase, and 68 hybridization reactions were performed according to the strategy outlined above. Hybridization signals due to repetitive sequences were eliminated by prehybridizing the probes to a high C_0t value with a mixture of cloned human repetitive sequences and human genomic DNA (11). Mixed probes detected a minimum of 9 and a maximum of 46 cross-hybridizing unique clones on the filter matrix with each hybridization reaction using a pooled probe (Fig. 3). To aid in the analysis of the data generated by this procedure, the *y* and *x* coordinates of the cross-hybridizing clones were recorded, and matches were identified by using computer programs specifically written for the analysis of these data. From this initial series of experiments, 1099 pairs of linked clones were detected from the hybridization of 36 pooled columns and 32 pooled rows of the matrix. Several of these predicted overlapping clones were analyzed by detailed restriction mapping and individual analysis of overlaps using end-specific RNA probes to confirm the predicted linkage.

From the list of linked clones produced by the initial multiplex analysis, contigs were assembled either manually or through computer analysis of the data from the predicted hybridization linkage using mixed multiple RNA probes. By using a simple tree-building algorithm for constructing contigs from the list of linked clones, 315 contigs were assembled from the 1099 linked clones detected from the initial multiplex analysis. The size of the predicted contigs ranged from 2 linked cosmids to 27 cosmids grouped into a contig potentially extending over 300 kilobases, with the majority of contigs consisting of between 2 and 5 linked cosmids. To confirm that these groupings reflected the true structure of the human chromosome and not artifactual groupings due to random cross-hybridization, several of the contigs were restriction mapped in detail to confirm the degree of physical overlap and to establish a physical map. The results of the

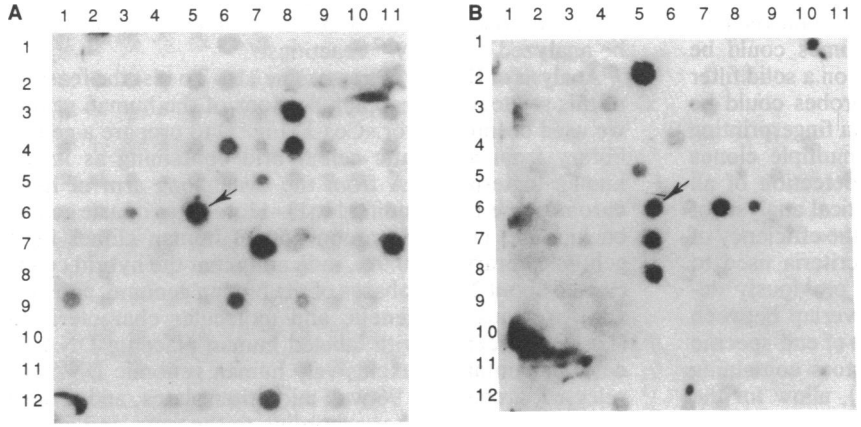

FIG. 3. Cosmid multiplex analysis of a collection of cosmids mapping to the long arm of human chromosome 11. (*A*) Multiplex analysis of human cosmid clones arrayed in a 36 × 36 matrix and hybridized with a mixed probe consisting of RNA transcripts from a clone of a row of the matrix. A portion of the filter is shown. (*B*) A portion of the filter shown in *A* hybridized with a mixed probe representing a pool of all cosmids aligned along a column of the matrix. The arrow identifies a cosmid clone that hybridizes with both mixed probes and is linked to the clone located at the intersection of the row and column from which probe mixtures were prepared.

multiplex analysis and restriction map of the resulting contig assembled by this strategy are shown in Fig. 4.

Completeness of the Multiplex Data. The cosmid set analyzed using T7 polymerase reactions on pooled clones represents about 1.5 genome equivalents, where the minimal detectable overlap can be as small as 200 nucleotides (10, 11, 15). The parameter θ has been defined as the fraction of length of two clones that must be shared in order for overlap to be detected (17). At $\theta = 0.05$, a maximum of about 450 contigs would be expected to result after the analysis of 1 genome equivalent, and about 260 contigs would be expected to result after the analysis of 2 genome equivalents (17). This analysis of clones representing about 1.5 genome equivalents, generating 315 contigs, is therefore in good agreement with theoretical predictions based on this value of θ. In addition, this strategy for multiplex analysis results in a theoretical loss of a small amount of the data since clones that are "linked" to a clone that by chance is present on a template row or column will be masked by hybridization signals of templates. Thus, the collection may contain some clones that overlap but are not detectable by the current strategy. In spite of this, a notable advantage of the current strategy is that the analysis of this number of clones proceeds according to the square root of the number of clones and a large amount of overlap data can be obtained with far less actual manipulation than if each clone were analyzed individually. The analysis of a number of clones sufficient to give a 10-fold redundancy using a similar strategy would be expected to result in far more complete ordered cosmid set.

Artifacts. The detection of false overlaps can be attributed to technical or theoretical factors. The initial analysis of genomic DNA by rapid chromosome "walking" (10, 15) suggested that repetitive DNA sequences lying near the ends of genomic cosmid inserts may result in artifactual linkage and present a major obstacle to the type of analysis proposed here. A simple prehybridization procedure where RNA probes are saturated with cloned human repetitive sequences was sufficient for reducing most of repetitive signal (11) and eliminating the vast majority of artifactual linkages. A second type of false linkage might result from the analysis using a single polymerase and the chance occurrence of one clone that is linked to two other cosmids, one present in a template row and one in a template column of the matrix. Thus, these clones would appear to overlap because of their common linkage to a third clone but would represent a false signal. Several examples of this type of artifactual linkage have been detected in this data set. This type of artifact occurs very rarely in a clone collection of this redundancy and, when the analysis is carried to completion using both T3 and T7 mixed RNA probes, the data generated would be internally redundant, which allows for the elimination of this type of artifact. Finally, it is anticipated that further refinement of this approach, perhaps through the use of more complex redundant pooling strategies, should eliminate most serious artifacts arising during multiplex clone analysis. Additional ordering of cosmid contigs and map closure may be carried out by *in situ* hybridization to metaphase or prophase chromosomes by using fluorescent-labeled cosmid clones (24) or by using linking clones derived from yeast artificial chromosomes (7).

DISCUSSION

The distal long arm of human chromosome 11 is the location of many genes of biological and medical significance. Chromosome 11q23 contains a cluster of genes encoding proteins that are members of the immunoglobulin superfamily and possibly important for cell–cell interactions in the immune and nervous systems, including Thy-1, CD3γ, -δ, and -ε, and the neural cell adhesion molecule (N-CAM) (25). Genes in which defects may be responsible for ataxia telangiectasia (26), multiple endocrine neoplasia type I (27), and human type I diabetes analogous to that of the nonobese diabetic (NOD) mouse (28) map to the long arm of chromosome 11, and translocation breakpoints at 11q13 and 11q23 are frequently found associated with acute leukemias, Ewing sarcoma, peripheral neuroepithelioma, and Askin tumor (29).

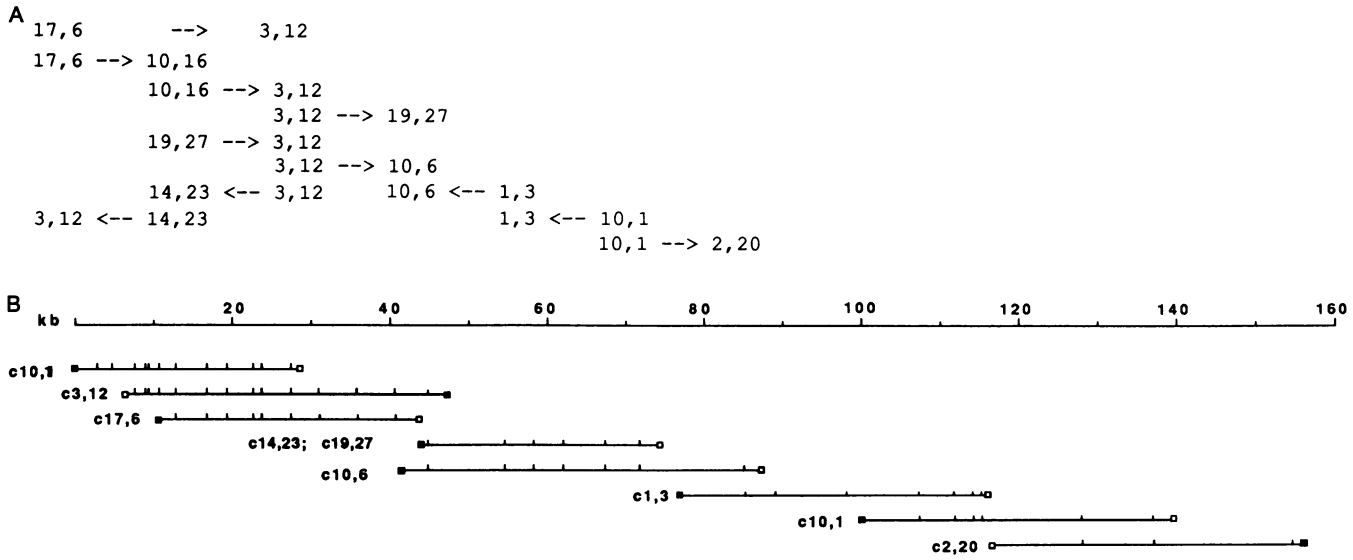

FIG. 4. Predicted contigs from human chromosome 11q and restriction enzyme digestion analysis. (*A*) Predicted linkage and orientation of a representative cosmid contig (contig no. 2) generated by multiplex analysis of the chromosome 11q cosmid set, detection of common cross-hybridizing clones, and assembly of a predicted contig by computer analysis. The computer output indicates the coordinates of linked clones (y, x), and the arrows denote the orientation of the linkage. For instance, 17,6 —> 10,16 indicates that the T7 end of the insert of the cosmid at coordinates $y = 17, x = 6$ overlaps with the cosmid at coordinates $y = 10, x = 16$. (*B*) Restriction map of the cosmid contig assembled in *A*. The restriction map of each of the overlapping clones was determined by the analysis of partial *Eco*RI digestion products hybridized with ^{32}P-labeled T3 or T7 promoter-specific oligonucleotides (11). Overlapping areas not confirmed by restriction map analysis were confirmed by hybridization analysis with end-specific RNA probes generated from individual cosmid clones. Cosmid clones c14,23 and c19,27 are identical. □, Bacteriophage T3 promoter; ■, bacteriophage T7 promoter; kb, kilobases.

This work represents an initial attempt to unravel the underlying biology of this genomic region.

We developed a strategy for rapid bottom-up cosmid mapping to expedite the analysis of this region of the genome. Multiplex analysis produces data that are analogous to clone fingerprinting (4, 5, 8) but that allow simultaneous analysis of pooled cosmids and therefore do not require heroic efforts. The use of hybridization rather than pattern matching decreases the minimal detectable overlap, θ, and thus reduces the number of clones needed for map closure by up to 3-fold (17). Furthermore, the analysis of cosmids in pools decreases the amount of effort required by an order of magnitude or more. In this pilot project, the actual manipulation of clones was reduced by 18-fold over procedures necessitating the individual analysis of cosmid clones. In addition, the strategy presented here is extremely simple and although biochemical reactions, data collection, and data analysis are all amenable to automation, the methodology can also be effectively applied without the use of expensive instrumentation.

In this regard, the present application of this strategy to a set of cosmids representing chromosome 11q has generated a set of cosmid contigs that includes most of the known gene loci and DNA markers and is estimated to include about 60% of the 11q13–11qter region. This degree of refinement was obtained by using only 68 analytical reactions and allows closure of gaps by using alternate methods. By using automated restriction mapping, the cosmid set has been analyzed for clones containing multiple rare restriction sites, which likely contain hypomethylated CpG-rich islands associated with many expressed genes (30), or for single rare sites, which are useful as linking clones for long range mapping using pulsed-field gel electrophoresis. Screening organized libraries with probes prepared from whole cDNA libraries may allow detection of cosmids containing genes that express abundant RNA transcripts. In addition, *in situ* hybridization to metaphase chromosomes has allowed ordering of many of the cosmid contigs on 11q (unpublished data). A more complete multiplex analysis based on a greater redundancy of cosmid clones, ensuring a 10-fold representation and utilizing both T3 and T7 promoters to generate internally consistent redundant data, would be expected to result in greater coverage and possibly in near closure of the map. This analysis would require a collection of about 10,000 cosmids and 200 T3 or T7 reactions/hybridizations rather than the 68 carried out here. Finally, the method proposed here represents a special case of a more general mapping strategy based on combinatoric pools of probes organized in matrices of n dimensions (G.A.E. and K. C. Evans, unpublished results) and is potentially applicable to sets of cosmids or yeast artificial chromosomes. With ongoing efforts to analyze complex mammalian genomes in great detail, these strategies for the assembly of cosmids or yeast artificial chromosomes into contigs may prove useful.

We would like to thank B. Rothenberg, J. Zhao, J. Eubanks, G. Andreason, W. Goad, E. Hildebrand, and S. Chen for discussions; D. Housman, T. Glazer, and M. Litt for exchange of reagents, cell lines, and information; J. Longmire, T. Friedmann, and R. Moyzis for plasmids; K. Sirotkin for advice on computer simulation; K. C. Evans for mathematical insight; G. Hermanson, D. McElligott, K. Pischel, and C. Landel for careful review of the manuscript; Stratagene for reagents; W. Gilbert for valuable suggestions; and R. Dulbecco for a continuing interest in this project. This work was supported in part by grants from the National Institutes of Health, Department of Energy, March of Dimes Birth Defects Foundation, and Hereditary Disease Foundation and by funds from the G. Harold and Leila Y. Mathers Charitable Foundation. G.A.E. is a Pew Scholar in the Biomedical Sciences supported by the Pew Memorial Trust.

1. Schwartz, D. C. & Cantor, C. R. (1984) *Cell* **37**, 67–75.
2. Southern, E. M., Anand, R., Brown, W. R. A. & Flecther, D. S. (1987) *Nucleic Acids Res.* **15**, 5925–5943.
3. Burke, D. T., Carle, F. G. & Olson, M. M. (1987) *Science* **236**, 806–812.
4. Olson, M. V., Dutchik, J. E., Graham, M. Y., Brodeur, G. M., Helms, C., Frank, M., MacCollin, M., Scheinman, R. & Frank, T. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7826–7830.
5. Coulson, A., Sulston, J., Brenner, S. & Karn, J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7821–7825.
6. Staden, R. (1980) *Nucleic Acids Res.* **8**, 3673–3694.
7. Coulson, A., Waterson, R., Kiff, J., Sulston, J. & Kohara, Y. (1988) *Nature (London)* **335**, 184–186.
8. Kohara, Y., Akiyama, K. & Isono, K. (1987) *Cell* **50**, 495–508.
9. Poustka, A., Pohl, T., Barlow, D. P., Zehetner, G., Craig, A., Michiels, F., Ehrlich, E., Frischauf, A. M. & Lehrach, H. (1986) *Cold Spring Harbor Symp. Quant. Biol.* **51**, 131–139.
10. Wahl, G. M., Lewis, K. A., Ruiz, J. C., Rothenberg, B. E., Zhao, J. & Evans, G. A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2160–2164.
11. Evans, G. A., Lewis, K. A. & Rothenberg, B. E. (1989) *Gene*, in press.
12. Pyati, J., Kucherlapati, R. S. & Skoultchi, A. I. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3435–3439.
13. Maslen, C. L., Jones, C., Glaser, T., Magenis, R. E., Sheey, R., Kellog, J. & Litt, M. (1988) *Genomics* **2**, 66–75.
14. Evans, G. A. & Wahl, G. M. (1987) *Methods Enzymol.* **152**, 604–610.
15. Evans, G. A., Lewis, K. A. & Lawless, G. M. (1988) *Immunogenetics* **28**, 365–373.
16. Vogeli, G. & Kaytes, P. S. (1987) *Methods Enzymol.* **152**, 407–415.
17. Lander, E. S. & Waterman, M. S. (1988) *Genomics* **2**, 231–239.
18. van Rijs, J., Giguere, V., Hurst, J., van Agthoven, T., ven Kessel, A. G., Goyert, S. & Grosfeld, F. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 5832–5835.
19. Watson, D. K., McWilliams-Smith, M. J., Koczak, C., Reeves, R., Gearhart, J., Nunn, M. F., Nash, W., Fowle, J. R., Duesberg, P., Papas, T. S. & O'Brien, S. J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1792–1796.
20. Wang, A. L., Arredondo-Vega, F. X., Giampietro, P. F., Smith, M., Anderson, W. F. & Desnick, R. J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 5734–5738.
21. Misrahi, M., Atger, M., d'Auriol, L., Loosfelt, H., Meriel, C., Fridlansky, F., Guiochon-Mantel, A., Galibert, F. & Milgrom, E. (1987) *Biochem. Biophys. Res. Commun.* **143**, 740–748.
22. Lauffer, L., Garcia, P. D., Harkins, R. N., Coussens, L., Ullrich, A. & Walter, P. (1985) *Nature (London)* **318**, 334–338.
23. Karathanasis, S. K., Zannis, V. I. & Breslow, J. L. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6147–6151.
24. Lichter, P., Cremer, T., Tang, C. J., Watkins, P. C., Manuelidis, L. & Ward, D. C. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 9664–9668.
25. Nguyen, C., Mattei, M. G., Mattei, J. F., Santoni, M. J., Goridis, C. & Jordan, B. R. (1986) *J. Cell Biol.* **102**, 711–715.
26. Gatti, R. A., Berkel, I., Boder, E., Braedt, G., Charmley, P., Concannon, P., Ersoy, F., Foroud, T., Jaspers, N. G. J., Lange, K., Lathrop, G. M., Leppert, M., Nakamura, Y., O'Connell, P., Paterson, M., Salser, W., Sanal, O. Silver, J., Sparkes, R. S., Susi, E., Weeks, D. E., Wei, S., White, R. & Yoder, F. (1988) *Nature (London)* **336**, 577–580.
27. Larsson, C., Skogseid, B., Oberg, K., Nakamura, Y. & Nordenskjold, M. (1988) *Nature (London)* **332**, 85–87.
28. Prochazka, M., Leiter, E. H., Serreze, D. V. & Coleman, D. L. (1987) *Science* **237**, 286–289.
29. Griffin, C. A., McKeon, C., Israel, M., Gegonne, A., Ghysdael, J., Stehelin, D., Douglass, E. C., Green, A. A. & Emanuel, B. S. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6122–6126.
30. Bird, A. (1986) *Nature (London)* **321**, 209–213.