Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 6218–6222, August 1990
Genetics

Physical mapping of human chromosomes by repetitive sequence fingerprinting

(cosmids/contigs/flow cytometry/chromosome 16)

RAYMOND L. STALLINGS, DAVID C. TORNEY, CARL E. HILDEBRAND, JONATHAN L. LONGMIRE, LARRY L. DEAVEN, JAMES H. JETT, NORMAN A. DOGGETT, AND ROBERT K. MOYZIS*

Center for Human Genome Studies, Los Alamos National Laboratory, University of California, Los Alamos, NM 87545

Communicated by Theodore T. Puck, May 18, 1990

ABSTRACT **We have developed an approach for identifying overlapping cosmid clones by exploiting the high density of repetitive sequences in complex genomes. Individual clones are fingerprinted, using a combination of restriction enzyme digestions followed by hybridization with selected classes of repetitive sequences. This "repeat fingerprinting" technique allows small regions of clone overlap (10–20%) to be unambiguously assigned. We demonstrate the utility of this approach, using the fingerprinting of 3145 cosmid clones (1.25× coverage), containing one or more $(GT)_n$ repeats, from human chromosome 16. A statistical analysis was used to link these clones into 460 contiguous sequences (contigs), averaging 106 kilobases (kb) in length and representing approximately 54% (48.7 Mb) of the euchromatic arms of this chromosome. These values are consistent with theoretical calculations and indicate that 150- to 200-kb contigs can be generated with 1.5× coverage. This strategy requires the fingerprinting of approximately one-fourth as many cosmids as random strategies requiring 50% minimum overlap for overlap detection. By "nucleating" at specific regions in the human genome, and exploiting the high density of interspersed sequences, this approach allows (*i*) the rapid generation of large (>100-kb) contigs in the early stages of contig mapping and (*ii*) the production of a contig map with useful landmarks for rapid integration of the genetic and physical maps.**

During the past several years, interest has increased in mapping the entire human genome (1). This interest has been catalyzed by innovations in molecular biology and recombinant DNA techniques (2–5) and the continuing search for efficient ways to isolate disease genes (6) and understand genome function (7, 8). The pioneering physical mapping projects on *Escherichia coli* (9–11), yeast (12), and *Caenorhabditis elegans* (13) DNA demonstrated the feasibility of using these techniques on genomes of small to moderate complexity. The human genome is 30 times larger than the *C. elegans* genome (13). However, individual human chromosomes have similar sequence complexities (5) and should present a similar challenge. While yeast artificial chromosomes (YACs) provide an efficient way to order large genomic regions (4), emphasis has been placed on the construction of ordered cosmid clone physical maps. Unlike YACs, such maps provide easily accessible materials for determining the biology of genome organization and for obtaining the DNA sequence, the ultimate physical map (1).

Early methods for obtaining ordered clone maps utilized a variety of strategies to identify overlaps between randomly isolated clones (10–15). In general, these approaches require a high degree of overlap between "fingerprints" to unambiguously determine linkage. As discussed by Lander and Waterman (16), the requirement for a large value of minimum detectable fractional clone overlap, or θ, inevitably leads to slow progress in the generation of contiguous sequence (contig) maps and the production of maps with low connectivity (i.e., numerous small contigs). Recent approaches to derive more information from clone fingerprints, and hence reduce θ, have employed (*i*) a combination of restriction fragment fingerprinting with the acquisition of additional DNA sequence information (17); and (*ii*) the use of a "multiplex" hybridization strategy, in which pools of probes obtained from the ends of individual clones are hybridized to arrayed sets of clones (18).

A potential problem for contig mapping of higher eukaryotic DNA is the high abundance of sequences present in multiple copies (19). Twenty-five percent of the human genome consists of repetitive DNA (20). Previous "low-resolution" maps of the distribution of these repetitive sequences indicated they contained sufficient information to aid contig mapping (20). Rather than being an impediment to physical mapping, the distributions of the three most abundant sequences, *Alu* (21), L1 (22), and $(GT)_n$ (23, 24), provide highly informative "tags" to fingerprint cloned genomic DNAs. We present a general strategy for contig fingerprinting using cosmid clones that reduces θ to an experimentally acceptable value (0.10–0.20). Further, we do not randomly fingerprint clones, but "nucleate," or pre-select clones known to contain a given repetitive sequence. Using this approach, one rapidly generates large [>100-kilobase (kb)] contigs in the early stages of physical mapping that cover a significant portion of the target genome (>50%). Equally as important, and unlike all previous contig mapping strategies (10–18), our approach generates contigs that (*i*) have known degrees of overlap of their constituent clones; (*ii*) can be reprobed with other sequences without further DNA isolation; and (*iii*) provide immediately accessible information on the distribution of both restriction enzyme sites and repetitive and nonrepetitive sequences in the mapped regions.

MATERIALS AND METHODS

***In Situ* Hybridization.** Methods described previously were used (7, 20).

Construction of a Chromosome-16-Specific Cosmid Library. Human chromosomes 16 were isolated from a somatic cell hybrid (CY18; ref. 26) by bivariate fluorescence-activated flow sorting (5). A single chromosome 16 was the only human chromosome present in this hybrid and was easily resolved from a mouse background. After partial digestion with *Sau*3AI and dephosphorylation with calf intestinal alkaline phosphatase, 336 ng of sorted chromosomal DNA was ligated to 0.5 μg of cloning arms from the cosmid vector sCos1 (27).

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

Abbreviation: YAC, yeast artificial chromosome.

*To whom reprint requests should be addressed.

In vitro packaging and infection of *E. coli* host strain HB101 yielded 1.75×10^5 independent recombinants, giving a 67-fold statistical representation. Twenty-six thousand independent clones (a 10-fold representation) from the primary library were selected as individual colonies and archived in 96-well microtiter dishes. This archival system allows rapid access to clones found to be positive after hybridization to any probe.

Isolation of Cosmid DNA and Automated Restriction Digestions. Cosmid DNA was isolated from 5-ml overnight cultures by a boiling procedure similar to that described by Holmes and Quigley (25). Restriction enzyme digestions were performed in 96-well microtiter dishes with reagents added by a Beckman Biomek 1000 automated work station. The restriction buffer used in each of these digestions was 80 mM NaCl/100 mM Tris·HCl, pH 7.5/5 mM $MgCl_2$ containing acetylated bovine serum albumin at 100 μg/ml. This buffer is compatible with both *Eco*RI and *Hin*dIII. Plates were covered with a Mylar sealer and incubated at 37°C overnight in a humidified chamber.

Electrophoresis and Southern Blotting. The digested cosmid DNA samples were loaded onto 20 × 25 cm 0.6% agarose gels. For size markers, *Hin*dIII- and *Eag* I-digested λ phage DNA plus *Hae* III-digested ϕX174 phage DNA were loaded in every seventh lane. Electrophoresis was in a horizontal apparatus (BRL model H1) for 40 hr at 27 V in 1× TA buffer (0.04 M Tris acetate). Southern blot transfers were performed as previously described (7).

Procedures for Constructing Hybridization Arrays of Bacterial Colonies on Membrane Filters. Cosmid clones were transferred from 96-well microtiter dishes onto beds of agarose with a stamping device containing 96 flat-bottomed prongs precisely matching the microtiter array. Imprints from 6 titer dishes (576 clones) were made on an agarose bed in a 22 × 22 cm plate. Two Grunstein–Hogness (28) colony lifts were made from each plate containing the arrayed bacterial colonies, using NEN GeneScreen*Plus* nylon hybridization membranes.

Probe Preparation and Hybridization. $(GT)_{25}$, $(AC)_{25}$, as well as consensus sequence oligodeoxynucleotides of *Alu* (20), L1 (22), α-satellite (29), β-satellite (30), and satellites I, II, and III (31) repetitive DNAs were synthesized on an Applied Biosystems 380B DNA synthesizer and labeled as described previously (7, 20, 29). Hybridization and washing conditions were described previously (7, 32).

Analysis of Negatives of Ethidium Bromide-Stained Gels and Autoradiograms. Polaroid negatives of ethidium bromide-stained gels were scanned with a BioImage Visage 110. The Whole Band Analysis software supplied by BioImage was used to determine molecular weights of restriction enzyme fragments. Southern hybridizations were scored on a console screen. Fingerprint data were transferred electronically to the Laboratory Notebook (developed at Los Alamos) data base.

Fingerprint Analysis. The Bayes formula is used to calculate the probability of overlap for two clone inserts:

$$P(\text{overlap}|\mathbf{S}) = p(\mathbf{S} \text{ and overlap})/p(\mathbf{S}). \quad [1]$$

In Eq. 1, S equals $\{S_E, S_H, S_{EH}\}$, a vector of three statistics, the components computed from both clones' *Eco*RI, *Hin*dIII, and double-digest fingerprints, since the latter are not independent. Eqs. 2 and 3 expand a quantity in Eq. 1, assuming both inserts have length $\bar{L}$.

$$p(\mathbf{S} \text{ and overlap}) = \int_0^{\bar{L}} dl\, p(\mathbf{S}|\text{overlap } l)\rho(\text{overlap } l). \quad [2]$$

$$\rho(\text{overlap } l) \propto 1 - 2\exp-[\bar{L}/\bar{l}_{GT}] + \exp-[(2\bar{L} - l)/\bar{l}_{GT}]. \quad [3]$$

Eq. 3 reflects the condition that clones accepted for fingerprinting contain at least one $(GT)_n$ sequence, "GT nucleation," under the premise that these sequences are randomly placed with average spacing $\bar{l}_{GT}$. Since $\bar{l}_{GT}$ is 40 kb, the average $\bar{L}$ is 37 kb, and the euchromatic part of chromosome 16 is 8.5×10^7 bases, integrating ρ(overlap l) gives an *a priori* overlap probability of 1.1×10^{-3}. Eqs. 1–3 suffice for the determination of insert overlap probabilities, given a suitable statistic S.

Each component S of S approximates the likelihood ratio of the data and overlap to the data and nonoverlap for one digest. The ratio is derived for a statistical model in reasonable agreement with the data. One begins with a matrix C with matrix elements:

$$c_{ij} \equiv \frac{H_{GT} \cdot H_{RS} \cdot \bar{l}_r \cdot \exp[(l_{1i} + l_{2j})/2\bar{l}_r] \cdot \exp-[(l_{1i} - l_{2j})^2/2\varepsilon^2(l_{1i}^2 + l_{2j}^2)]}{\varepsilon\sqrt{2\pi(l_{1i}^2 + l_{2j}^2)}}. \quad [4]$$

In Eq. 4, l_{1i} is the length of restriction fragment from the first clone, l_{2j} is the length of restriction fragment j from the second clone, $\bar{l}_r$ is the average length between restriction sites, and ε times the length of a fragment is the standard deviation of length measurement reproducibility; ε equals 0.5%. Also, H_{GT} and H_{RS} are factors reflecting results of hybridization to GT repetitive sequence and Cot1 repetitive sequence probes. These H are a function of λ, the ratio of the average length of compared fragments to the average distance between occurrences of the corresponding hybridization site. If both fragments hybridize, H is $\exp(\lambda)$; if neither fragment hybridizes, H is $[1 - \exp(\lambda)]^{-1}$; otherwise H is 0. Naturally, most c_{ij} are negligible.

S is derived from C as follows:

$$S = \sum_{k=1}^{\min(N_1, N_2)} \sigma_k, \quad [5]$$

where

$$\sigma_k = \frac{(N_1 - k)!(N_2 - k)!}{N_1!N_2!} \sum_{i_1, i_2, \ldots, i_k = 1}^{N_1}{}' \sum_{j_1, j_2, \ldots, j_k = 1}^{N_2}{}' \prod_{l=1}^{k} c_{i_l j_l};\ N_1, N_2 \geq k. \quad [6]$$

N_1 and N_2 are the number of fragments of the two inserts. The primes on the summation signs in Eq. 6 indicate that no two indexes are equal. Programs TRUE and FALSE compute p(S) conditioned on overlap and nonoverlap. Contigs are assembled by linking all insert pairs with overlap probabilities greater than 0.9.

RESULTS

Rationale for Mapping Using Repeat Sequences. All eukaryotic genomes contain DNA sequences present in higher than expected abundance (19). These repetitive DNA sequences can be tagged with synthesized oligomer consensus sequences. In any eukaryotic genome, certain classes of interspersed repetitive sequences will have an optimal abundance for targeting. For example, the sequence $(GT)_n$, where $n \approx$ 20–25, is a ubiquitous component of eukaryotic genomes, occurring once on average every 50–100 kb (20, 23, 24). For cosmid-sized fragments (40 kb) cloned from human DNA, a synthetic $(GT)_{25}$ sequence should hybridize to approximately 50% of the recombinants. Consensus oligomers to various regions of L1 repeats, because of their 5′ truncation, target

human DNA at different frequencies (22). Extreme 5′ L1 oligomers hybridize once every 300–400 kb in human DNA (R.L.S. and R.K.M., unpublished results), providing optimal targeting for YAC-sized DNA fragments.

If the targeting frequency of a given repetitive sequence is coupled with additional fingerprint information, a high degree of information is obtained. Restriction fragment fingerprints are useful in this regard. By superimposing information about the molecular weights of DNA obtained from a given clone with information about whether each fragment does or does not contain a given repetitive DNA sequence, a unique fingerprint can be obtained. For example, approximately 1500–2500 $(GT)_n$ sequences are present on human chromosome 16 (approximately 95 Mb in length). For cosmid clones obtained from this chromosome, a 20-cm agarose gel is capable, at a minimum, of resolving 100 distinct restriction fragments with sizes between 0.5 and 25 kb. If a single $(GT)_n$ site is present on a given clone, then three separate restriction enzyme digests, with enzymes recognizing 6 bp, followed by Southern blot analysis for the presence or absence of $(GT)_n$ sequences, yields 1 of potentially $100 \times 100 \times 100$ or 10^6 different GT-positive restriction patterns. Two cosmid clones from a chromosome-specific recombinant DNA library with the same pattern of GT hybridization have an extremely high probability of overlap, even if they share only the GT-positive region (Fig. 1). Detailed algorithms were derived to estimate this probability, as outlined in *Materials and Methods*.

Sequential hybridization of restriction digests of individual chromosome-specific cosmid clones with GT, *Alu* (20), and L1 (22) consensus oligomers yields a unique fingerprint, with the minimum detectable overlap, θ, being 0.10–0.20. This range of θ will significantly accelerate the construction of large (>100-kb) contigs in the early stages of physical mapping, in comparison to strategies requiring θ values of 0.5 (16). To further accelerate mapping, a nonrandom strategy can be employed. Since the goal of all long-range mapping projects is to obtain high connectivity of the entire region of interest, the starting point is irrelevant. Therefore, preselecting clones which contain a given repetitive sequence "nucleates" mapping at these sites, guaranteeing that the minimum θ will be obtained. Further, the rate and length of initial contig generation are accelerated, proportional to the extent of clustering of the original nucleation sites. For example, preselecting $(GT)_n$-containing clones nucleates mapping at 1500–2500 sites on chromosome 16. *In situ* hybridization analysis suggests that GT sites are uniformly distributed over

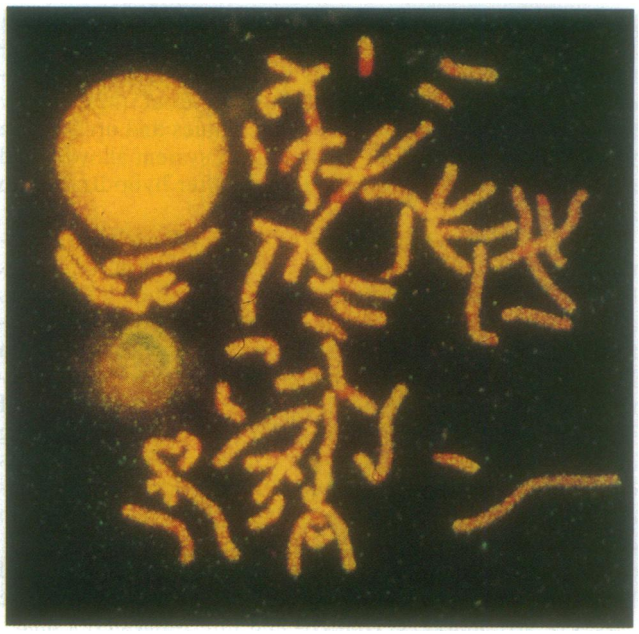

FIG. 2. *In situ* hybridization analysis of human chromosomes with biotin-labeled $(AC)_{25}$ (yellow) as probe. GT sequences are underrepresented at the centromeric regions of some chromosomes and the distal half of Yq. However, GT sequences appear to be more uniformly distributed on all euchromatic chromosome regions, unlike *Alu* repeats (20).

euchromatic arms of all human chromosomes (Fig. 2). We have focused on a particular application of this general repetitive sequence fingerprinting procedure to initiate mapping of human chromosome 16.

Library Characterization. A chromosome-16-specific cosmid library was constructed and a 5× representation (1× ≈ 2600) was screened for the presence of sequences from single-copy genes, multigene families, and anonymous probes. All sequences were found in numbers proportional to their expected abundances: 58 MT, 15 HP, 4 D16S4, 3 CTRB, 2 TAT, 2 LCAT, and 1 APRT clones. From this analysis, one can conclude that the representation of DNA sequences in this library is sufficient to initiate the construction of a global contig map of chromosome 16.

Prescreening of the Library with Repetitive Sequence Probes for Ordered Clone Mapping. A subset (7500 clones) of the

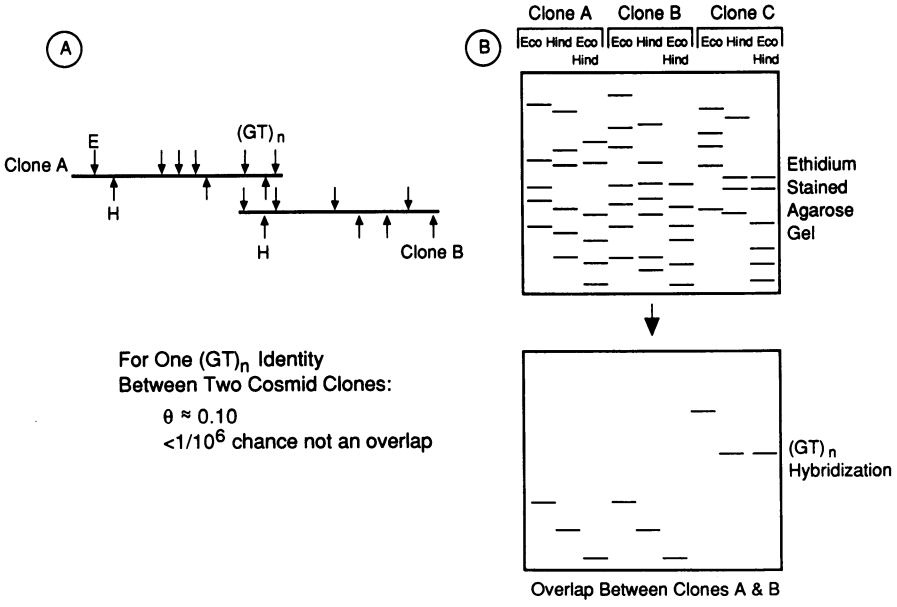

FIG. 1. Rationale behind the repeat fingerprinting process. (*A*) Two clones, A and B, overlap by approximately 10%. However, a $(GT)_n$ repeat sequence is shared by the two clones in the overlap region, and hence information on restriction fragment size coupled with $(GT)_n$ hybridization yields an extremely high probability of overlap. ↓, *Eco*RI sites; ↑, *Hin*dIII sites. (*B*) Diagram of an ethidium bromide-stained gel and a Southern blot of the gel hybridized to the $(GT)_n$ repeat probe. Clones A and B share a few restriction fragments of the same size and have identical $(GT)_n$-positive restriction fragments. Clone C does not overlap A and B. Overlap between clones A and B can be detected on the basis of the $(GT)_n$ hybridization information alone.

26,000 clones that were arrayed and archived in microtiter dishes were replicated onto nylon hybridization membranes (576 clones per membrane). Clones containing GT repeats, centromeric satellite repeats, and mouse DNA inserts were identified. Mouse inserts result from small amounts (approximately 7%) of contaminating mouse chromosomes during flow sorting. Following hybridization with a set of pooled consensus oligomers for five human centromeric tandem repeats [α-satellite (29), β-satellite (30), and satellites I, II, and III (31)], it was determined that approximately 4.5% of the clones were composed of satellite DNA. Clones containing tandem repeats were catalogued for future analysis, as constructing ordered clone maps from the centromeric regions of human chromosomes is not necessary at present.

All 13 grids, containing 7500 clones, were hybridized to labeled GT sequence probes. Forty-one percent of the clones were GT positive in this colony assay, and clones that were negative for mouse insert or satellite DNA inserts were used in our approach to ordered clone mapping. Analysis of a small fraction of GT-negative clones indicated that approximately 62% of all cosmid clones contain at least one GT repeat, and some of these are missed by colony screening. This number is consistent with the estimate that 1500–2500 $(GT)_n$ sites are present on chromosome 16 (20).

Progress in Large-Scale Mapping Using the Repetitive Sequence Fingerprinting Approach. Fingerprint analysis was performed on a total of 3145 cosmid clones. Of these clones, 2692 were GT-positive. Among the GT-negative clones fingerprinted, 57 were preselected on the basis of being positive for the 5′ region of the L1 repeat. Fig. 3 shows a representative gel. Single digestions with *Eco*RI and *Hin*dIII, as well as an *Eco*RI/*Hin*dIII double digestion, were conducted

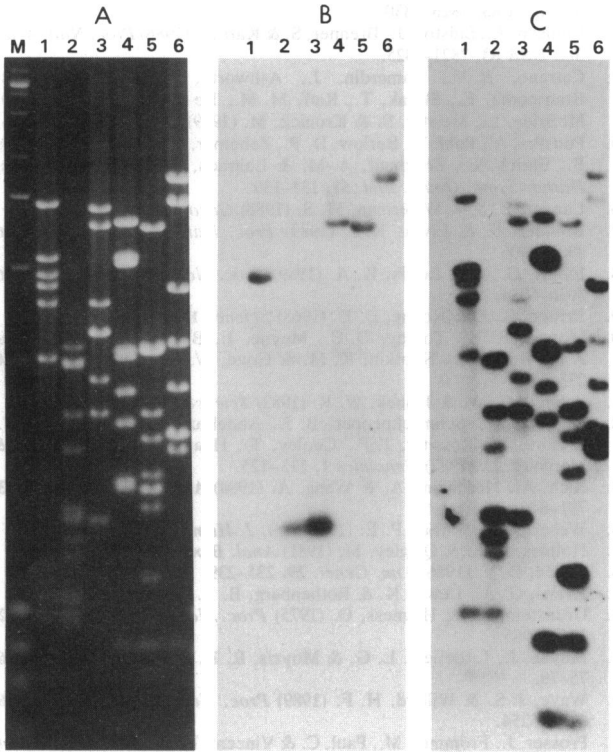

FIG. 3. Examples of autoradiograms of blots from ethidium bromide-stained gels (*A*) hybridized to labeled GT (*B*), and Cot1-fractionated (*C*) DNA. Data from two cosmid clones, 26C7 and 26C8, are shown. Lanes 1, 2, and 3 contain DNA from cosmid 26C7 digested with *Eco*RI, *Eco*RI/*Hin*dIII, and *Hin*dIII, respectively, while lanes 4, 5, and 6 contain DNA from cosmid 26C8 digested with the same series of enzymes, respectively. In *A*, lane M contains size standards.

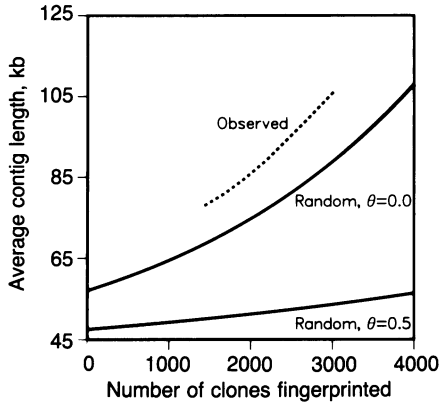

FIG. 4. Observed average contig length (two or more overlapping clones) is plotted as a function of number of fingerprinted cosmid clones (dotted curve). Theoretical predictions of average contig length (kb) are plotted for random strategies detecting minimal overlap ($\theta = 0.0$) and detecting >50% overlap ($\theta = 0.5$) (solid curves). These predictions are generated by equations derived from ref. 16. Contigs generated by repeat fingerprinting and GT nucleation are substantially larger than theoretically predicted for random strategies, including those detecting minimal overlap. Note that equations from ref. 16 pertain to islands (contigs and singletons) and further derivations were necessary to apply these equations to contigs.

for each clone. Southern blot analysis was conducted by using standard procedures (7). The molecular weight of each restriction fragment was determined and data were analyzed by computational algorithms described in *Materials and Methods*. Fifty-four percent of the clones contained a single GT-positive restriction fragment (Fig. 3*B*), while 30%, 11%, 4%, and 1% contained 2, 3, 4, and 5 or more GT-positive fragments, respectively. This observed distribution is consistent with a random distribution of $(GT)_n$ sequences along chromosome arms (Fig. 2). In this initial analysis, fragments containing Cot1 hybridizing sequences (predominantly *Alu*, L1, and GT repeats) were also determined (Fig. 3*C*). Experiments using *Alu* or L1 consensus oligomers rather than Cot1 DNA have yielded equally informative fingerprinting information (data not shown). A total of 2823 pairs of overlapping clones emerged from this analysis. Using a second algorithm, we assembled these 2823 overlapping cosmid clones into 460 contigs. The mean overlap between cosmids was 43% and the minimum overlap detected was 10%. The increase in contig size with the number of clones fingerprinted is plotted in Fig. 4. A typical large contig containing 23 overlapping clones spanning 240 kb is shown in Fig. 5. The average contig size was 106 kb (the average number of cosmids per contig is 4.6). These contigs represent 54% of the euchromatic arms of human chromosome 16. Using end cosmids as probes, we have independently confirmed 12 out of 12 contigs tested by pulsed-field gel analysis (data not shown).

DISCUSSION

We have developed an approach for identifying overlapping cosmid clones by exploiting the high density of repetitive sequences in complex genomes. By coupling restriction digestion mapping (12) with oligomer probes targeting abundant interspersed repetitive sequences (20), a "fingerprint" is obtained with sufficient information content to unambiguously link two cosmid-sized clones obtained from a chromosome-specific DNA library (Fig. 1). This approach expands on previous pioneering work to exploit repetitive sequences for genetic mapping (33). The technique is generally applicable to any genome, simply requiring a minimum knowledge of the overall distribution of repetitive sequences in the target genome (20). A similar approach based on

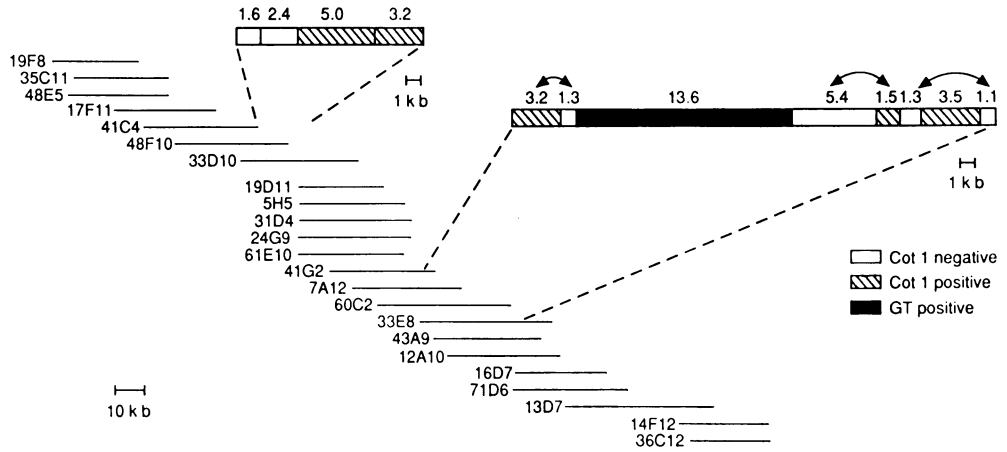

FIG. 5. An example of a 240-kb contig (no. 33) obtained after the analysis of GT-positive cosmid clones. The expanded maps at the top show that some ordering of restriction fragments (*Eco*RI in this case) can be obtained based on differences in overlap. Information as to whether the fragments contain $(GT)_n$ sequences, other repeat sequences, or single-copy DNA is also obtained.

hybridization of *Alu* and L1 repeats has been implemented for fingerprinting human YACs (34).

The utility of this repeat fingerprinting procedure has been demonstrated by analyzing 3145 cosmid clones from chromosome 16; 460 contigs were obtained, averaging 106 kb in length and representing 54% of the nonsatellite regions of this chromosome (Figs. 2, 4, and 5). A modest extrapolation reveals that contigs 150–200 kb in length can be generated by using this method with only 1.5× coverage of the euchromatic arms. To obtain comparably large contigs from a random strategy requiring 50% overlap for detection would require 4 times the number of fingerprinted cosmids (Fig. 4). Fingerprinting 4000 vs. 16,000 clones is a substantial reduction in effort. The larger size of the contigs generated by this approach is an effect of nucleation and a reduced θ, and it may indicate clustering of some $(GT)_n$ sites.

The ability to rapidly generate "YAC-sized" (4) contigs directly from cosmid clones (Figs. 4 and 5) allows immediate access to large regions of the human genome. Such contigs provide the materials to immediately generate information on human genome organization and accelerate the isolation of disease genes. They provide the materials to continue contig growth and ordering, using a variety of techniques (4), including "renucleation" at other repetitive DNA sites. Once the initial contigs are constructed, many of these activities can be done distributively, either by direct distribution of grided filter arrays (5) or through the use of a sequenced tagged site (STS) language (35). The ability to generate 100- to 200-kb contigs, encompassing half of a chromosome (Figs. 4 and 5), coupled with pulsed-field mapping of "gap" size (2, 3), provides the materials for the immediate development of STS markers with a spacing of 100 kb on chromosome 16. The highly polymorphic $(GT)_n$ sequences (24) used to generate these contigs are of sufficient density to generate a high-resolution (<1 centimorgan) genetic map, allowing direct integration between the genetic and physical maps of human chromosomes.

We acknowledge the technical contributions of Sarah Anzick, Nancy Brown, Amanda Ford, Linda Meincke, Chris Munk, Summers Cox, Liz Saunders, Judy Tesmer, and Lynn Clark. The IBM/Los Alamos National Laboratory Joint Study provided an IBM 3090 computer. We thank Thomas Marr, Debra Nelson, Michael Cannon, Rebecca Koskela, Doyce Nix, Steven White, David Balding, Randall Dougherty, and Karen Schenk for assistance with analysis and Julie Meyne for providing the *in situ* hybridization photo. We also acknowledge Grant Sutherland, Michael Karin, John McClean, N. Maeda, G. Scherer, and J. Tischfield for the use of the D16S4, MT, LCAT, HP, TAT, and APRT probes, respectively. This work was supported by the U.S. Department of Energy (Grants B04836/F698, B04860/F116, B04861/F118, and B04718/F518).

1. McKusick, V. A. & Ruddle, F. H. (1987) *Genomics* **1**, 103–106.
2. Schwartz, D. C. & Cantor, C. R. (1984) *Cell* **37**, 67–75.
3. Cantor, C. R., Smith, C. L. & Mathew, M. K. (1988) *Annu. Rev. Biophys. Biophys. Chem.* **17**, 287–304.
4. Burke, D. T., Carle, G. F. & Olson, M. V. (1987) *Science* **236**, 806–812.
5. Deaven, L. L., Van Dilla, M. A., Bartholdi, M. F., Carrano, A. V., Cram, L. S., Fuscoe, J. C., Gray, J. W., Hildebrand, C. E., Moyzis, R. K. & Perlman, J. (1986) *Cold Spring Harbor Symp. Quant. Biol.* **51**, 159–167.
6. Orkin, S. H. (1986) *Cell* **47**, 845–850.
7. Moyzis, R. K., Buckingham, J. M., Cram, L. S., Dani, M., Deaven, L. L., Jones, M. D., Meyne, J., Ratliff, R. L. & Wu, J.-R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6622–6626.
8. Stallings, R. L., Munk, A. C., Longmire, J. L., Jett, J. H., Wilder, M. E., Siciliano, M. J., Adair, G. M. & Crawford, B. D. (1985) *Chromosoma* **92**, 156–163.
9. Smith, C. L., Econome, J. G., Schutt, A., Klco, S. & Cantor, C. R. (1987) *Science* **236**, 1448–1453.
10. Daniels, D. L. & Blattner, F. R. (1987) *Nature (London)* **325**, 831–832.
11. Kohara, Y., Akiyama, K. & Isono, K. (1987) *Cell* **50**, 495–508.
12. Olson, M. V., Dutchik, J. E., Graham, M. V., Brodeur, G. M., Frank, M., MacColin, M., Scheinman, R. & Frank, T. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7826–7830.
13. Coulson, A., Sulston, J., Brenner, S. & Karn, J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7821–7825.
14. Carrano, A. V., Lamerdin, J., Ashworth, L. K., Watkins, B., Branscomb, E., Slezak, T., Raff, M. M., De Jong, P. J., Keith, D., McBride, L., Meister, S. & Kronick, M. (1989) *Genomics* **4**, 129–136.
15. Poustka, A., Pohl, T., Barlow, D. P., Zehetner, G., Craig, A., Michiels, F., Ehrich, E., Frischauf, A.-M. & Lehrach, H. (1986) *Cold Spring Harbor Symp. Quant. Biol.* **51**, 131–139.
16. Lander, E. S. & Waterman, M. S. (1988) *Genomics* **2**, 231–239.
17. Brenner, S. & Livak, K. J. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8902–8906.
18. Evans, G. A. & Lewis, K. A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5030–5034.
19. Britten, R. J. & Kohne, D. E. (1968) *Science* **161**, 529–540.
20. Moyzis, R. K., Torney, D. C., Meyne, J., Buckingham, J. M., Wu, J. R., Burks, C., Sirotkin, K. M. & Goad, W. B. (1989) *Genomics* **4**, 273–289.
21. Schmid, C. W. & Jelinek, W. R. (1982) *Science* **215**, 1065–1070.
22. Scott, A. F., Schmeckpepper, B. J., Abdelrazik, M., Comey, C. T., O'Hara, B., Rossiter, J. P., Cooley, T., Heath, P., Smith, K. D. & Margolet, L. (1987) *Genomics* **1**, 113–125.
23. Rich, A., Nordheim, A. & Wang, A. (1984) *Annu. Rev. Biochem.* **53**, 791–846.
24. Weber, J. L. & May, P. E. (1989) *Am. J. Hum. Genet.* **44**, 388–396.
25. Holmes, D. S. & Quigley, M. (1981) *Anal. Biochem.* **114**, 193–197.
26. Callen, D. F. (1986) *Ann. Genet.* **29**, 235–239.
27. Evans, G. A., Lewis, K. & Rothenberg, B. E. (1989) *Gene* **79**, 9–20.
28. Grunstein, M. & Hogness, O. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961–3965.
29. Meyne, J., Littlefield, L. G. & Moyzis, R. K. (1989) *Mutat. Res.* **226**, 75–79.
30. Waye, J. S. & Willard, H. F. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6250–6254.
31. Prosser, J., Frommer, M., Paul, C. & Vincent, P. C. (1986) *J. Mol. Biol.* **187**, 145–155.
32. Riethman, H. C., Moyzis, R. K., Meyne, J., Burke, D. T. & Olson, M. V. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6240–6244.
33. Gusella, J. F., Jones, C., Kao, F.-T., Housman, D. & Puck, T. T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7804–7808.
34. Wada, M., Little, R. D., Abidi, F., Porta, G., Labella, T., Cooper, T., Della Valle, G., D'Vruso, M. & Schlessinger, D. (1990) *Am. J. Hum. Genet.* **46**, 95–106.
35. Olson, M., Hood, L., Cantor, C. & Botstein, D. (1989) *Science* **245**, 1334–1335.