Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 3149–3154, April 1996
Genetics

A high-resolution physical map of human chromosome 11

SHIZHEN QIN*†, NORMA J. NOWAK*†‡, JIALU ZHANG*, SHEILA N. J. SAIT*, PETER G. MAYERS*§, MICHAEL J. HIGGINS*, YI-JUN CHENG*, LI LI*, DAVID J. MUNROE¶, DANIELA S. GERHARD‖, BERNHARD H. WEBER**, EVA BRIC¶, DAVID E. HOUSMAN¶, GLEN A. EVANS††, AND THOMAS B. SHOWS*‡‡

*Department of Human Genetics, Roswell Park Cancer Institute, Buffalo, NY 14263; ¶Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139; ‖Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110; **Julius Maximilians Universitate Wurzberg, Institute for Humangenetik, D-97074 Wurzberg, Germany; and ††The Eugene McDermott Center for Growth and Development, University of Texas Southwestern Medical Center, Dallas, TX 75235

Contributed by David E. Housman, December 14, 1995

ABSTRACT **The development of a highly reliable physical map with landmark sites spaced an average of 100 kbp apart has been a central goal of the Human Genome Project. We have approached the physical mapping of human chromosome 11 with this goal as a primary target. We have focused on strategies that would utilize yeast artificial chromosome (YAC) technology, thus permitting long-range coverage of hundreds of kilobases of genomic DNA, yet we sought to minimize the ambiguities inherent in the use of this technology, particularly the occurrence of chimeric genomic DNA clones. This was achieved through the development of a chromosome 11-specific YAC library from a human somatic cell hybrid line that has retained chromosome 11 as its sole human component. To maximize the efficiency of YAC contig assembly and extension, we have employed an *Alu*-PCR-based hybridization screening system. This system eliminates many of the more costly and time-consuming steps associated with sequence tagged site content mapping such as sequencing, primer production, and hierarchical screening, resulting in greater efficiency with increased throughput and reduced cost. Using these approaches, we have achieved YAC coverage for >90% of human chromosome 11, with an average intermarker distance of <100 kbp. Cytogenetic localization has been determined for each contig by fluorescent *in situ* hybridization and/or sequence tagged site content. The YAC contigs that we have generated should provide a robust framework to move forward to sequence-ready templates for the sequencing efforts of the Human Genome Project as well as more focused positional cloning on chromosome 11.**

High-fidelity physical maps of each chromosome will facilitate the sequencing efforts of the Human Genome Project as well as the identification and localization of human disease genes. Construction of such maps has been simplified by recent technological advances such as yeast artificial chromosome (YAC) cloning and the widespread use of PCR-based screening systems for arrayed libraries (1). Application of these methods has resulted in the construction of low-order physical maps, in the form of YAC contigs, for chromosome 21q (2) and the euchromatic region of the Y chromosome (3). These contigs were ordered and developed largely on the basis of sequence tagged site (STS) content. Assembly of these contigs was facilitated by prior knowledge of STS order across the target regions, obtained, in the case of the 21q map, by a set of well-characterized chromosome 21-specific somatic cell hybrid mapping panels integrated with a dense set of ordered genetic markers (2). Similarly, a large collection of naturally occurring Y chromosome breaks, used in conjunction with a Y chromosome-enriched YAC library, were vital to the rapid development of the Y map (3). The production of YAC contigs spanning other chromosomes or chromosome arms has proved to be more difficult. The difficulties encountered with the development of such maps can largely be attributed to (*i*) the comparative lack of similar ordered mapping reagents available for other chromosomes, (*ii*) the relative inefficiency of STS-content mapping, and (*iii*) the inherent physical and technical limitations of whole genome YAC library screening including their large size and high rate of chimerism. We sought to directly address these limitations during the development of a YAC contig-based physical map of chromosome 11. In an effort to minimize many of the ambiguities associated with the screening of whole genome YAC libraries, we have developed an arrayed chromosome 11-specific YAC library from a somatic cell hybrid line that has retained chromosome 11 as its sole human component (4). The small size and essentially nonchimeric nature of this library has accelerated contig assembly and greatly increased the sensitivity of screening in comparison to that of whole genome libraries. In addition, as an alternative to STS-content mapping, we have employed an *Alu*-PCR-based hybridization system for the assembly of large YAC clone contigs (5, 6). This system offers several advantages over STS-content mapping with respect to increased throughput, efficiency, and cost reduction (5, 6). As a result, we have achieved YAC coverage for ≈130 Mbp, or >90% of chromosome 11, with an average intermarker distance of <100 kbp. Furthermore, since each of the 1824 clones in the library has been sized and they are largely devoid of chimeras, an accurate assessment of intermarker distance can be estimated from these contigs.

MATERIALS AND METHODS

YAC Libraries. The 4X chromosome 11-specific YAC library was prepared from the J1 monochromosomal hybrid, screened against hamster Cot-1 DNA to eliminate interspecies chimeras, and arrayed into 19 96-well microtiter plates as described (4). Each YAC has been assigned an address based upon its location within a plate, row, and column. All 1824 clones have been sized by pulsed-field gel electrophoresis. The average insert size is 350 kb. The CEPH mega YAC library was constructed, arrayed, and characterized as described (7).

YAC Library Pooling Schemes and DNA Preparation. The chromosome 11-specific YAC library was arranged into three blocks of six microtiter plates each. Individual YACs from within each block were grown to saturation and combined into a series of pools corresponding to rows, columns, and "half-

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.

Abbreviations: FISH, fluorescent *in situ* hybridization; STS, sequence tagged site; YAC, yeast artificial chromosome.

†S.Q. and N.J.N. contributed equally to this work.
‡For questions regarding the data base and resources, direct e-mail to nowak@shows.med.buffalo.edu.
§For questions regarding the server, direct e-mail to mayers@shows.med.buffalo.edu.
‡‡To whom reprint requests should be addressed.

plates." Row, column and half-plate pools contained 24, 24, and 48 individual YAC clones, respectively. High-purity DNAs were prepared from each of the 182 pools in agarose plugs by the lyticase/LiCl dodecyl sulfate method (7). The CEPH mega YAC library was arranged into pools as described (7).

***Alu*-PCR Amplification of YAC DNAs.** *Alu*-PCR amplification from all templates was directed from the *Alu* S/*Alu* J, *Alu*-end, and 47-23 primer sets (8). *Alu*-PCR amplification of YAC DNA pools was performed in a 100-μl reaction mixture containing 10 mM Tris (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM each dNTP, and 2.5 units of AmpliTaq DNA polymerase (Perkin–Elmer). *Alu*-PCR amplifications from the *Alu*-end primer were performed with 3.5 mM $MgCl_2$. The thermocycling parameters used were 94°C for 1 min, 58°C for 1 min, and 72°C for 45 sec for 35 cycles followed by 72°C for 4 min.

Preparation and Screening of *Alu*-PCR Hybridization Membranes. YAC DNA pools were *Alu*-PCR amplified with each individual *Alu* primer set. *Alu*-PCR amplification products from each pool were visually inspected on ethidium bromide-stained agarose gels and immobilized onto 8 × 12 cm nylon membranes using a manual offset spotting device (John Krietler, Washington University machine shop, Washington University, St. Louis). Filters were processed by baking for 1–2 hr at 80°C, denaturation in 0.4 M NaOH/0.5 M NaCl for 10 min, and neutralization in 0.5 M Tris, pH 8.0/0.5 M NaCl for 5 min. Alu-PCR product probes for hybridization were generated from cosmid, phage, or YAC template DNAs as described above. The entire set of PCR amplification products corresponding to each template was ethanol precipitated and labeled by random priming. Probes were preannealed with an equal volume of human placental DNA (10 μg/ml) and 0.3 volume of 1 M sodium phosphate (pH 8.0) for 2 hr at 65°C. Nylon filters were prehybridized for 2–18 hr at 42°C in 5× Denhardt's solution, 5× SSC, 0.1% SDS, 50% formamide, and salmon sperm DNA at 100 mg/ml. After overnight hybridization at 42°C, filters were rinsed twice in 2× SSC at room temperature for 5 min and then washed twice in 0.1× SSC/0.5% SDS at 65°C for 30 min. Exposure times varied between 2 hr and 2 days.

STS Screening. The chromosome 11-specific YAC library was screened for a total of 278 STSs. Established STSs were obtained either through the Genome Data Base (Baltimore) or as described in Smith *et al.* (9). In addition, several unique STSs were generated from YAC insert ends (10). All PCR reactions were done in a 15-μl reaction volume using a Perkin–Elmer/Cetus 9600 thermal cycler as described (9). All positives were verified by PCR using DNA prepared from individual clones.

Fluorescent *in Situ* Hybridization (FISH) Mapping. FISH analysis was performed as described (4).

Data Analysis and Contig Assembly. Contigs were assembled using SEGMAP (10), an interactive graphical tool for analyzing and displaying physical mapping data. The follow-

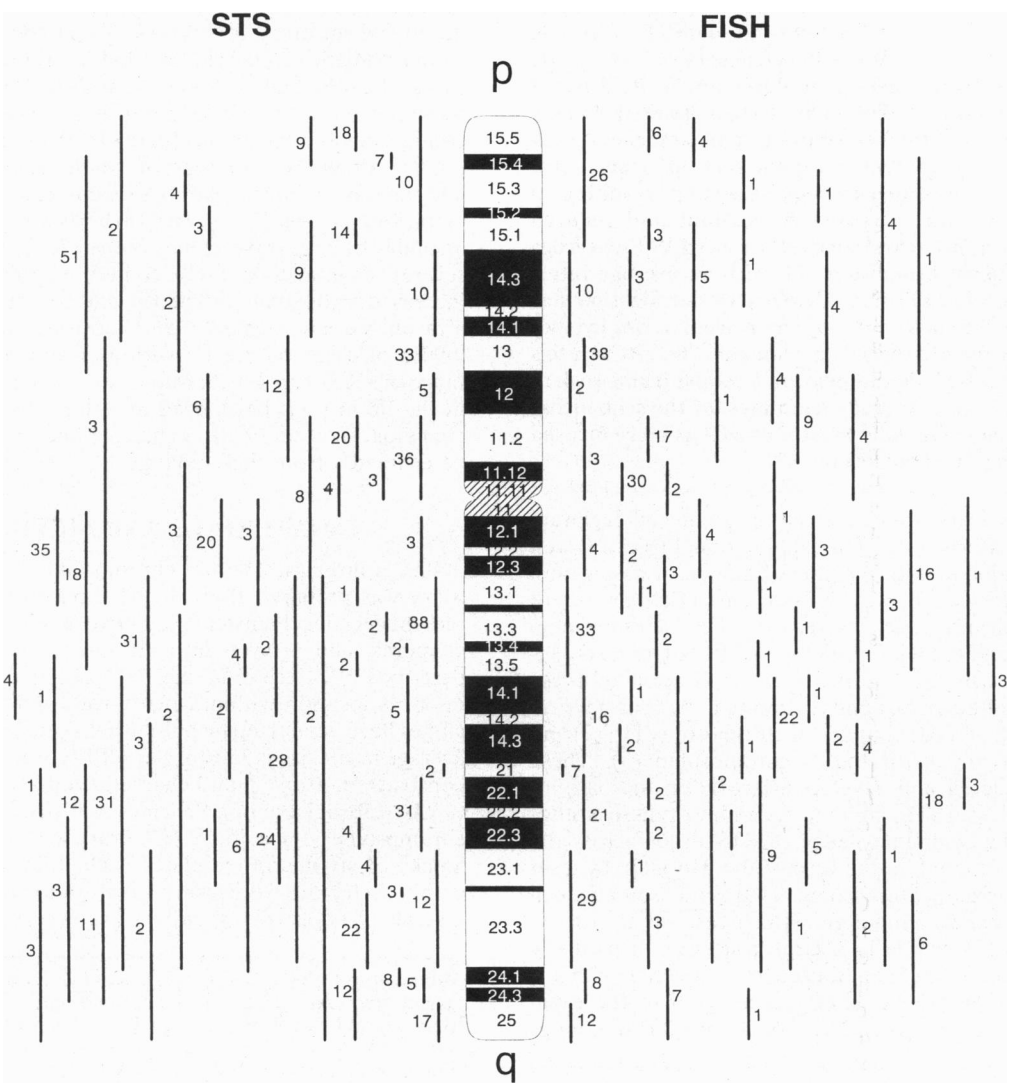

FIG. 1. Approximately 500 YAC clones were localized to specific bins indicated by the vertical bars on chromosome 11 by STS content (*Left*) and FISH (*Right*).

ing nomenclature has been used: (*i*) YAC clones from the chromosome-specific library have the prefix yRP followed by the plate address, (*ii*) YAC clones from the CEPH mega YAC collection have the prefix yMega followed by the plate address, (*iii*) *Alu*-PCR probes have the prefixes ySJ (S/J), y47 (47–23), and y3 (*Alu*-end) followed by the plate address of the YAC clone from which they were derived, (*iv*) STSs derived from the YAC clone insert ends are designated yRP or mega followed by the plate address with RE or LE for the right or left end of the insert, respectively, and (*v*) STSs derived from anonymous DNA sequences and genes have nomenclature assigned to them by the Genome Data Base.

Restriction Map Analysis of Contigs. Restriction enzyme digests of YAC DNAs prepared in agarose were carried out as recommended by the supplier (New England BioLabs); spermidine was added (to a final concentration of 5 mM) to buffers with >50 mM NaCl. Partial digests were achieved using serial dilutions of the enzyme with incubations ranging from 30 min to 4 hr. All reactions were allowed to equilibrate on ice for at least 1 hr prior to incubation at the appropriate temperature. Pulsed-field gel electrophoresis was carried out on the CHEF-DRII system (Bio-Rad). The DNA samples were analyzed on 1% agarose gels in 0.5× TBE (1× TBE is 89 mM Tris-borate/89 mM boric acid/2 mM EDTA, pH 8.0) at 200 V for 22 hr with ramping from 10 s to 50 s or 20 s to 60 s. Transfer of pulsed-field gels to nylon membranes and hybridization were as described (12). The membranes were sequentially hybridized with three sets of probes: (*i*) 2.6-kb and 1.7-kb fragments from pBR322 digested with *Bam*HI and *Pvu* II, which are homologous to the right and left arm of the pYAC4 vector, respectively; (*ii*) human Cot-1 DNA; or (*iii*) inter-*Alu* probes generated from the individual YAC clones.

RESULTS

Generation of Inter-*Alu* PCR Product Hybridization Probes. Inter-*Alu* PCR product hybridization probes were generated from individual YAC clones using *Alu*-specific primers. Six hundred fifty-four probes generated with primer S/J, 404 probes generated with primer 47–23, and 50 probes generated with the *Alu*-end primer were utilized in the final phase of this study. The low number of probes generated from the *Alu*-end primers reflects the degree of contig assembly that was already achieved by the time this primer set was used for hybridization and not its failure to generate successful hybridizing probes. Each YAC clone yielded from 4 to >10 PCR products when visualized on 1.5% agarose/ethidium bromide gels irrespective of the primer set. This includes YAC clones that had previously been mapped to Giemsa dark bands. The products ranged in size from <100 bp to >1 kb. Greater than 95% of the pooled products proved to be successful as probes. *Alu*-PCR products were also generated from a smaller set of chromosome 11-specific phage and cosmid clones.

Approximately 1100 inter-*Alu* probes, generated from individual YAC clones, were hybridized to filters stamped with inter-*Alu* products generated with the corresponding primer from YAC clone pools (see *Materials and Methods*). Screening ambiguities were resolved by examining half-plate pools or by Southern blot analysis of individual YAC clone inter-*Alu* PCR fingerprints.

STS Content Mapping. YAC clones have been identified for 278 STSs representing 62 genes, 171 anonymous DNA segments, and 45 YAC clone insert ends. On average, each PCR assay identified three or four individual YAC clones, as would be expected with a fourfold library. However, the STSs generally identified contigs previously assembled by *Alu*-PCR product hybridization, and only two STSs successfully joined separate contigs.

Contig Anchoring. Approximately 500 YAC clones (27% of the library) have been localized to specific bands on chromosome 11 by FISH (Fig. 1). An additional 200 clones (11% of the library) are anchored by virtue of containing a mapped STS (Fig. 1) (ref. 9; Genome Data Base). As a result, every contig contains at least one, and usually several, clones anchored by FISH and/or STS content. The localization data is presented both graphically and in a tabular (pter-qter) format (Figs. 2 and 3; WWW server URL, http://shows.med.buffalo.edu/home.html).

Verification of Contigs. Over 85% of the 119 contigs assembled by *Alu*-PCR product hybridization and STS content have been verified by Southern blot analysis of *Alu*-PCR fingerprints. These data have also served to confirm all single YAC linkages as well as resolve ambiguities inherent in the pooling scheme. Further verification of YAC contig integrity has been provided by pulsed-field restriction mapping. YAC clones comprising four contigs, chosen at random, have been subjected to pulsed-field restriction analysis. As demonstrated in Fig. 2, these restriction maps confirm contig integrity with respect to relative order and extent of overlap between clones.

Data Analysis. Analysis of both the hybridization and STS content data by SEGMAP has resulted in the assembly of 119 contigs ranging in size from 275 kb to 6100 kb and containing from 2 to 97 YAC clones, respectively, with 61 singletons (single YAC with at least one probe). An example of a map generated by SEGMAP for contig ySJ-1a2 is shown in Fig. 3. SEGMAP utilizes the YAC clone size data, determined by pulsed-field gradient analysis of each clone (depicted in parentheses below the clones), and also incorporates localization information either directly above the clones, if the localization was determined by FISH analysis, or above the STS contained in the clones for assembling the contig maps. Restriction map analysis of individual YAC clones making up the shortest tiling path through the contig demonstrates that the relative order of YAC clones, as well as intermarker distances predicted by SEGMAP, are essentially correct (Fig. 2). Thus, while this program was originally designed to analyze STS content data defining single points, it is also capable of handling complex mapping information generated by inter-*Alu* PCR product probes, which may represent the entire length of a YAC clone. The predicted size estimates of individual contigs are generally within 10% of the actual size.

Chromosome 11 Data Base. The data base can be searched using a YAC clone address or probe, and each of the corresponding contigs can be graphically viewed. Contig maps and tables containing all screened STSs and FISH localizations are available and can be searched either directly for a particular probe or YAC clone or in a pter-qter mode for regional information (http://shows.med.buffalo.edu/home.html).

DISCUSSION

A central goal of the Human Genome Project has been the development of a highly reliable physical map with landmark sites spaced an average of 100 kbp apart. The rationale for this goal is the ability to use such a map as a framework for the development of sequence-ready genomic DNA clone sets as well as the identification and cloning of human disease genes. We have approached the physical mapping of human chromosome 11 with this goal as a primary target. We focused on strategies that would utilize YAC technology, thus permitting long-range coverage of hundreds of kilobases of genomic DNA, yet sought to minimize the ambiguities inherent in the use of this technology, particularly the occurrence of chimeric genomic DNA clones. To achieve this goal, we developed a chromosome 11-specific YAC library from a human somatic cell hybrid line that has retained chromosome 11 as its sole human component (4). The reduced complexity and essentially nonchimeric nature of this library has translated into significantly increased efficiency and sensitivity of screening when compared to that of whole genome libraries. At the same time,

in an effort to maximize the efficiency of contig assembly and extension, we employed an *Alu*-PCR-based hybridization screening system (5, 6). This system eliminates many of the more costly and time-consuming steps associated with STS-content mapping such as sequencing, primer production, and hierarchical screening, resulting in greater efficiency with increased throughput and reduced cost (5, 6). Using these approaches, we have assembled the 1824 clone chromosome 11-specific YAC library into 119 contigs representing >90% of the chromosome with an average intermarker spacing of <100 kbp. Significantly, because the chromosome 11-specific YAC library is largely devoid of chimeric clones (4), sizing and

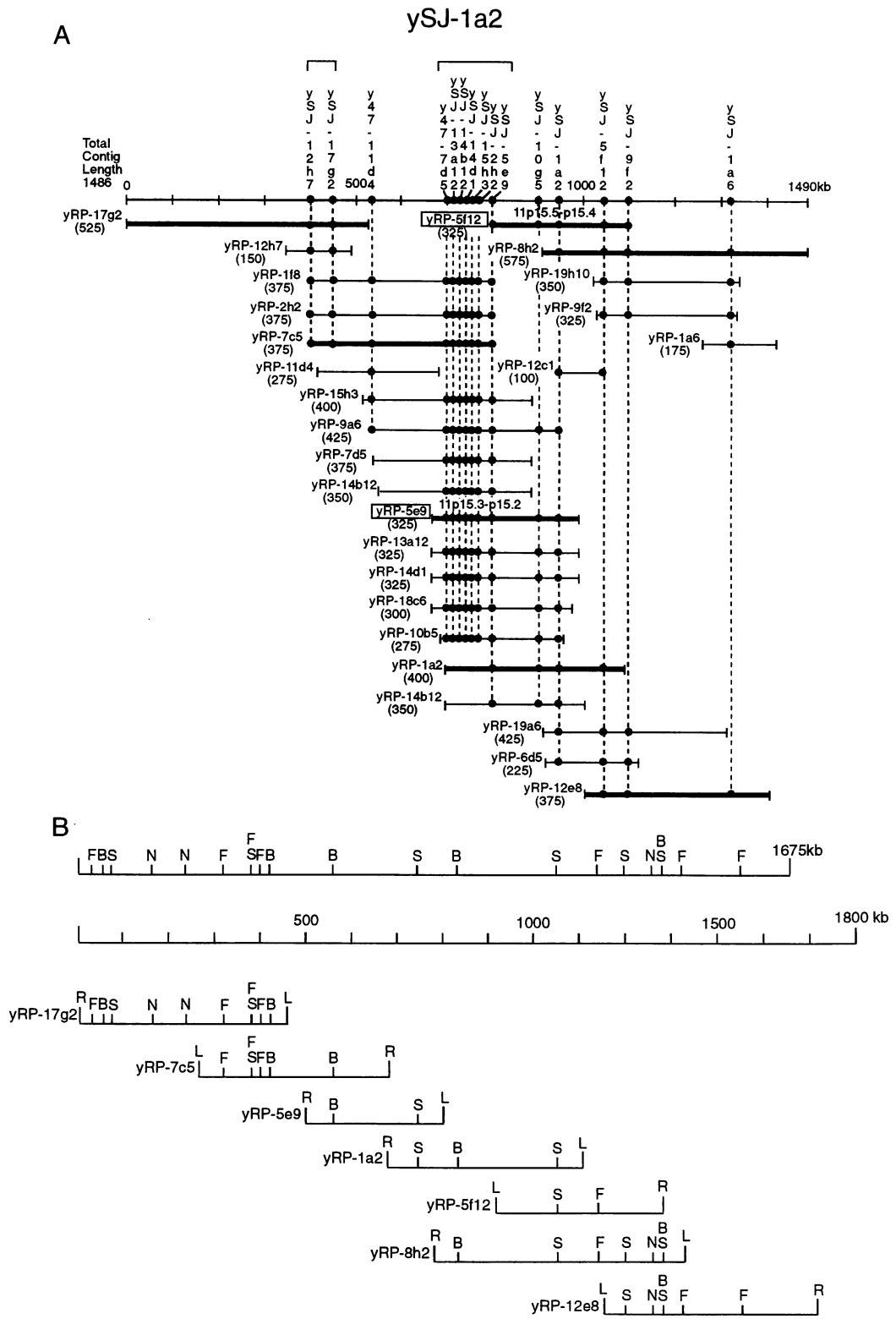

FIG. 2. Restriction map of contig ySJ-1a2 using YAC clones making up the shortest path through the contig. High molecular weight DNA from YAC clones (appearing in bold on the map) yRP-17g2, yRP-7c5, yRP-5e9, yRP-1a2, yRP-5f12, yRP-8h2, and yRP-12e8 was digested as described in *Materials and Methods* with *Not* I (N), *Sfi* I (F), *Sal* I (S), and *Bss*HII (B) and separated by electrophoresis on CHEF gels followed by Southern blot analysis. L and R represent the left and right ends of the clones, respectively.

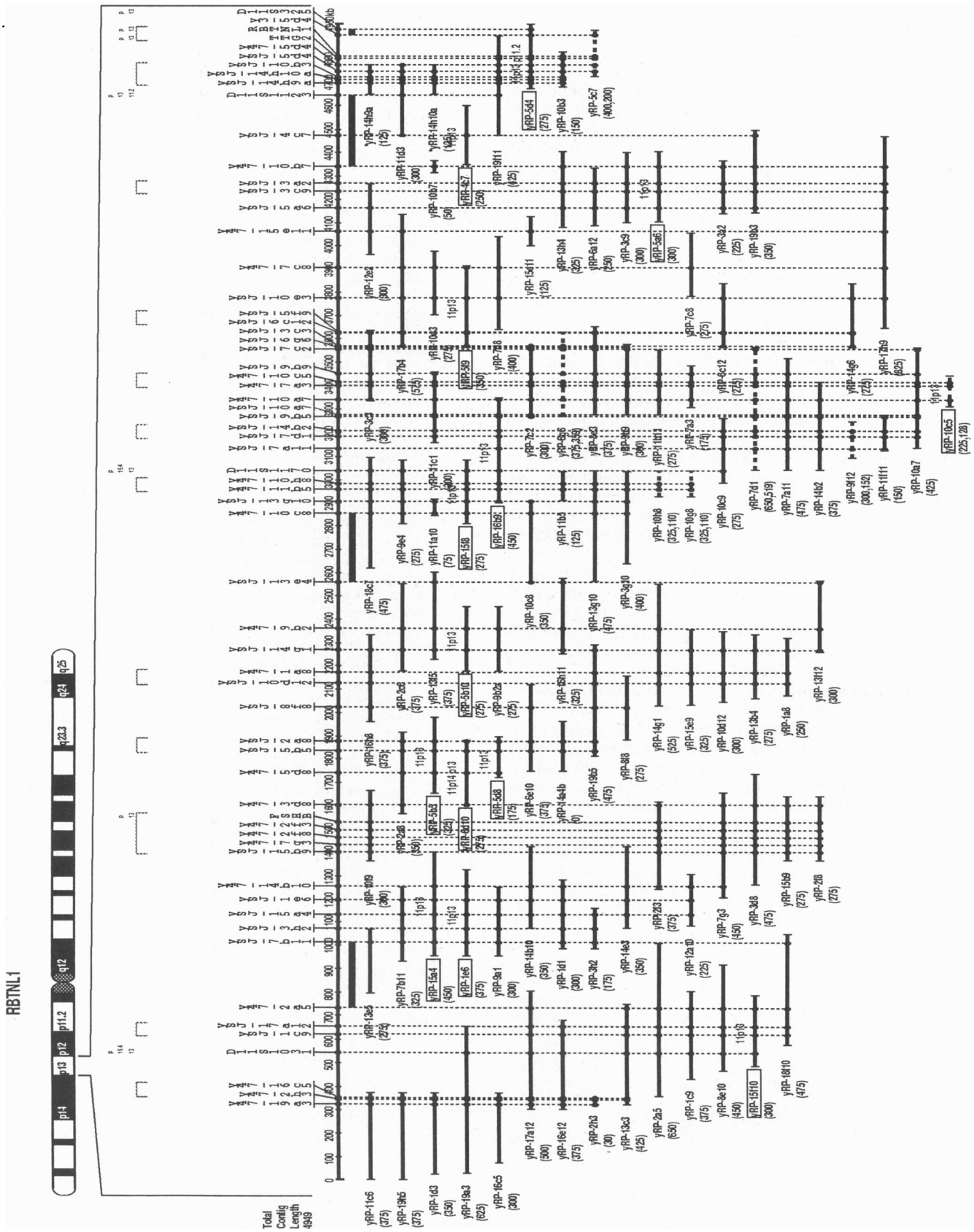

FIG. 3. A contig map of ySJ-1a2 generated by SEGMAP. Inter-*Alu* probes appear vertically above the clones making up the contigs. Sizes of individual YAC clones are denoted in parentheses beneath the clone address. FISH localization information appears above the clones. Localizations for mapped STSs appear vertically above the STS.

restriction mapping information garnered from these YAC contigs closely reflects that found in genomic DNAs (Fig. 2 and data not shown).

The average size of the assembled contigs (≈1 Mbp) is somewhat smaller than would be predicted (1–2 Mbp) considering the number of inter-Alu probes (≈1100) used to assemble the contigs (13). This difference may be a consequence of some degree of unevenness in the distribution of *Alu* elements in the genome (5, 6, 8). The density of *Alu* elements was clearly sufficient to support contig assembly by the *Alu*-PCR-based hybridization. However, clustering of *Alu* probes could lead to a significant degree of undetected overlap among the contigs.

The localization information for each of the contigs suggests the existence of several interesting structural features of chromosome 11 including duplications, low-order repetitive elements, chromosome-specific repetitive elements, and homologous regions on other chromosomes. Several YAC clones, from the chromosome-specific library, consistently mapped to two different regions on the chromosome by FISH analysis. CEPH mega YAC clones covering the same area similarly exhibit this dual localization. End clones and anonymous DNA segments from one such chromosome 11-specific YAC were sequenced, converted to STSs, and mapped by PCR on a chromosome 11-specific somatic cell hybrid mapping panel. These STSs mapped simultaneously to both the p and q arms of chromosome 11, suggesting the presence of an intrachromosomal duplication (data not shown). Approximately 10% of the chromosome-specific YAC clones map to two or more locations on chromosome 11, consistent with the presence of low-copy number, chromosome-specific, repetitive elements as has been suggested for chromosomes 5 and 7 (14, 15). Similarly, 15% of the clones also detected specific loci on other chromosomes, suggesting the presence of homologous regions or low-order repetitive elements.

The general principles that we have exploited in the construction of this physical map can easily be transferred to other chromosomes or chromosomal arms. The YAC contigs presented here should provide a robust framework to move forward to sequence-ready genomic templates as part of the sequencing efforts of the Human Genome Project as well as more focused positional cloning on chromosome 11.

We acknowledge the excellent technical assistance of Roger Eddy, Linda Haley, W. Michael Henry, and Stacey Simpson. This work was supported by National Institutes of Health Grants HG00359, HG00333, CA63333, and EY10514 and The Retinitis Pigmentosa Foundation (T.B.S.); National Cancer Institute Grant HL41486 (D.J.M.); American Cancer Society Grant CN64 and The Retinitis Pigmentosa Foundation (D.S.G.); National Institutes of Health Grant HG00299 (D.E.H.); National Center for Human Genome Research HG00102 and Department of Energy Grant DE EG03-88ER60694 (G.A.E.).

1. Green, E. D. & Olson, M. V. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 1213–1217.
2. Chumakov, I., Rigault, P., Guillou, S., *et al.* (1992) *Nature (London)* **359,** 380–387.
3. Foote, S., Vollrath, D., Hilton, A. & Page, D. C. (1992) *Science* **258,** 60–66.
4. Qin, S., Zhang, J., Isaacs, C., Nagafuchi, S., Sait, S. N. J., Abel, K., Higgins, M., Nowak, N. & Shows, T. B. (1993) *Genomics* **16,** 580–585.
5. Aburatani, H., Stanton, V. & Housman, D. E. (1996) *Proc. Natl. Acad. Sci. USA*, in press.
6. Liu, J., Stanton, V. P., Jr., Fujiwara, T. M., Wang, J.-X., Rezonzew, R., Crumley, M. J., Morgan, K., Gros, P., Housman, D. & Schurr, E. (1995) *Genomics* **26,** 178–191.
7. Anand, R., Riley, J. H., Butler, R., Smith, J. C. & Markham, A. F. (1990) *Nucleic Acids Res.* **18,** 1951–1956.
8. Munroe, D. J., Haas, M., Bric, E., Whitton, T., Aburatani, H., Hunter, K., Ward, D. & Housman, D. E. (1994) *Genomics* **19,** 506–514.
9. Smith, M. W., Clark, S. P., Hutchinson, J. S., Wei, Y. H., Churukian, A. C., Daniels, L. B., Diggle, K. L., Gen, M. W., Romo, A. J., Lin, Y., Selleri, L., McElligott, D. L. & Evans, G. A. (1993) *Genomics* **17,** 699–725.
10. Kere, J., Nagaraja, R., Mumm, S., Ciccodicola, A., D'Urso, M. & Schlessinger, D. (1992) *Genomics* **14,** 241–248.
11. Green, E. D. & Green, P. (1991) *PCR Methods Appl.* **1,** 77–90.
12. Higgins, M. J., Smilinich, N. J., Sait, S., Koenig, A., Pongratz, J., *et al.* (1994) *Genomics* **23,** 211–222.
13. Arratia, R., Lander, E. S., Tavare, S. & Waterman, M. S. (1991) *Genomics* **11,** 806–827.
14. Thompson, T., DiDonato, C., Simard, L., Ingraham, S., Burghes, A., Crawford, T., Tochette, C., Mendell, J. & Wasmuth, J. (1995) *Nat. Genet.* **9,** 56–62.
15. Kunz, J., Scherer, S. W., Klawitz, I., Soder, S., Du, Y.-Z., Speich, N., Kalff-Suske, M., Heng, H. H. Q., Tsui, L.-C. & Grzeschik, K.-H. (1994) *Genomics* **22,** 439–448.