tase, and the Eco RI fragment of pEC208 was inserted. Correctly oriented constructs were identified by restriction mapping of the resulting transformants with Hae III. Purified DNA was linearized by digestion with Sst I (which cuts once within the 3′ untranslated region of pEC208), and blunt ends were formed by treatment with the Klenow fragment of DNA polymerase I; this template was used for in vitro transcription by bacteriophage SP6 RNA polymerase [D. A. Melton, P. A. Krieg, M. R. Rebagliati, K. Zinn, M. R. Green, *Nucleic Acids Res.* **12**, 7035 (1984)]. The resulting RNA was translated in a rabbit reticulocyte lysate system (Promega Biotec) in the presence of [^{35}S]methionine (50 μCi) and immunoprecipitated with rabbit polyclonal antibodies to N-CAM (*23*).
38. M. R. Gazin, M. Rigolet, J.-P. Briand, M. H. V. Van Regenmortel, F. Galibert, *EMBO J.* **5**, 2241 (1986).
39. J. W. Tamkun *et al.*, *Cell* **46**, 271 (1986).
40. J. Lawler and R. O. Hynes, *J. Cell Biol.* **103**, 1635 (1986).
41. R. Yamamoto *et al.*, *Cell* **39**, 27 (1984).
42. S. Hoffman *et al.*, *J. Biol. Chem.* **257**, 7720 (1982).
43. G. J. Cole, A. Loewy, N. V. Cross, R. Akeson, L. Glaser, *J. Cell Biol.* **103**, 1739 (1986).
44. S. Hoffman and B. A. Cunningham, unpublished results.
45. M. Watanabe, A. L. Frelinger III, U. Rutishauser, *J. Cell Biol.* **103**, 1721 (1986).
46. G. M. Edelman, *Biochemistry* **9**, 3188 (1970).
47. L. M. Amzel and R. J. Poljak, *Annu. Rev. Biochem.* **48**, 961 (1979).
48. B. A. Cunningham, *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **35**, 1171 (1976).
49. J. W. Becker and G. N. Reeke, Jr., *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4225 (1985).
50. S. Hoffman, G. M. Edelman, B. A. Cunningham, unpublished results.
51. E. A. Kabat, T. T. Wu, H. Bilofsky, M. Reid-Miller, H. Perry, *Sequences of Proteins of Immunological Interest* (National Institutes of Health, Bethesda, MD, 1983).
52. A. F. Williams, A. N. Barclay, M. J. Clark, J. Gagnon, in *Gene Expression during Normal and Malignant Differentiation* (10th Sigrid Juselius Foundation Symposium), L. C. Andersson, C. G. Gahmberg, P. Ekblom, Eds. (Academic Press, New York, 1985), p. 125.
53. A. Bourgois, *Immunochemistry* **12**, 873 (1975).
54. A. D. McLachlan, *Protides Biol. Fluids Proc. Colloq.* **28**, 29 (1980).
55. T. Maddon and R. Axel, personal communication.
56. D. R. Littman and S. N. Gettner, *Nature (London)* **325**, 453 (1987).
57. M. Arquint *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 600 (1987).
58. J. L. Salzer, W. P. Holmes, D. R. Colman, *J. Cell Biol.*, **104**, 957 (1987).
59. J. G. Sutcliffe, R. J. Milner, T. M. Shinnick, F. E. Bloom, *Cell* **33**, 671 (1983).
60. A. F. Williams and J. Gagnon, *Science* **216**, 696 (1982).
61. I. F. Charo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8351 (1986).
62. E. Ruoslahti and M. D. Pierschbacher, *Cell* **44**, 517 (1986).
63. M. J. Humphries, S. K. Akiyama, M. Komoriya, K. Olden, K. M. Yamada, *J. Cell Biol.* **103**, 2637 (1986).
64. S. Hoffman, C.-M. Chuong, G. M. Edelman, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6881 (1984).
65. W. E. Gall and G. M. Edelman, *Science* **213**, 903 (1981).
66. E. G. Pollerberg, M. Schachner, J. Davoust, *Nature (London)* **324**, 462 (1986).
67. E. G. Pollerberg, R. Sadoul, C. Goridis, M. Schachner, *J. Cell Biol.* **101**, 1921 (1985).
68. G. Gennarini *et al.*, *J. Neurosci.* **6**, 1983 (1986).
69. G. M. Edelman, *Science* **192**, 218 (1976).
70. M. Barthels *et al.*, *EMBO J.* **6**, 907 (1987).
71. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
72. A. M. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1983).
73. R. M. K. Dale, B. A. McClure, J. P. Houchins, *Plasmid* **13**, 31 (1985).
74. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
75. D. W. Cleveland, S. G. Fischer, M. W. Kirschner, U. K. Laemmli, *J. Biol. Chem.* **252**, 1102 (1977).
76. D. J. Lipman and W. R. Pearson, *Science* **227**, 1435 (1985).
77. We thank V. Malik, E. Pichersky, and J. MacGregor for help with aspects of DNA sequencing; M. Burgoon for assistance with purification of the anti–N-CAM 1–reactive peptide; J. Salzer, D. Colman, J. Roder, and R. Axel, who communicated results prior to publication; and P. Ferrie, S. Lieberman, V. Sone, L. Tomchak, M. Petruzziello, B. Weerbrouck, and J. Diller for technical assistance. Supported by NIH grants HD-16550, HD-09635, AM-04256, and by a Senator Jacob Javits Center for Excellence in Neuroscience award, NS-22789, and a grant from Johnson and Johnson. Protein sequence analysis was carried out by the Rockefeller University Sequencing Facility, supported in part by the U.S. Army Research Office for the purchase of equipment.

Research Articles

Cloning of Large Segments of Exogenous DNA into Yeast by Means of Artificial Chromosome Vectors

DAVID T. BURKE, GEORGES F. CARLE, MAYNARD V. OLSON

Fragments of exogenous DNA that range in size up to several hundred kilobase pairs have been cloned into yeast by ligating them to vector sequences that allow their propagation as linear artificial chromosomes. Individual clones of yeast and human DNA that have been analyzed by pulsed-field gel electrophoresis appear to represent faithful replicas of the source DNA. The efficiency with which clones can be generated is high enough to allow the construction of comprehensive libraries from the genomes of higher organisms. By offering a tenfold increase in the size of the DNA molecules that can be cloned into a microbial host, this system addresses a major gap in existing experimental methods for analyzing complex DNA sources.

STANDARD RECOMBINANT DNA TECHNIQUES INVOLVE THE in vitro construction of small plasmid and viral chromosomes that can be transformed into host cells and clonally propagated. These cloning systems, whose capacities for exogenous DNA range up to 50 kilobase pairs (kb), are well suited to the analysis and manipulation of genes and small gene clusters from organisms in which the genetic information is tightly packed. It is increasingly apparent, however, that many of the functional genetic units in higher organisms span enormous tracts of DNA. For example, the bithorax locus in *Drosophila*, which participates in the regulation of the development of the fly's segmentation pattern, encompasses approximately 320 kb (*1*). The factor VIII gene in the human, which encodes the blood-clotting factor deficient in hemophilia A, spans at least 190 kb (*2*). Recent estimates of the size of the gene that is defective in Duchenne's muscular dystrophy suggest that this

The authors are in the Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110.

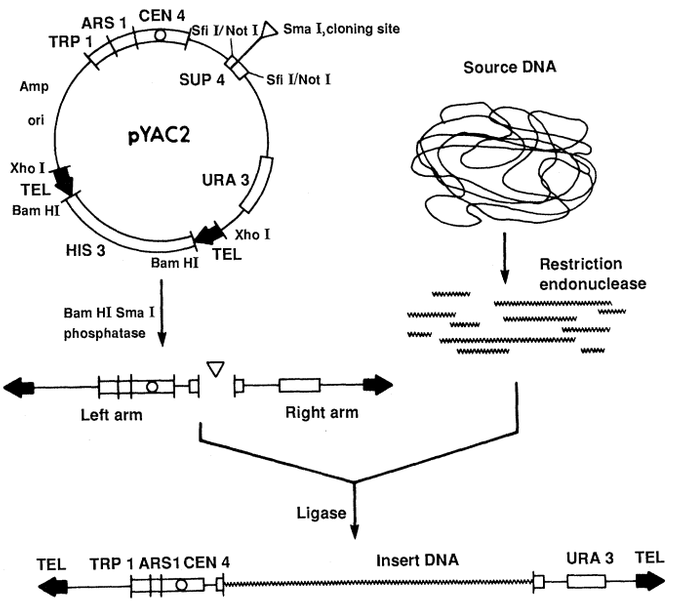

Fig. 1. Yeast artificial chromosome (YAC) cloning system. In the diagram of the vector pYAC2, pBR322-derived sequences are shown as a thin line. *SUP4, TRP1, HIS3*, and *URA3* are yeast genes: *SUP4* is an ochre-suppressing allele of a tyrosine transfer RNA gene that is interrupted when exogenous DNA is cloned into the vector; *TRP1* and *URA3* are present in the artificial chromosomes and allow selection for molecules that have acquired both chromosome arms from the vector; *HIS3* is discarded during the cloning process. *ARS1* and *CEN4* are sequences that are naturally adjacent to *TRP1* on yeast's chromosome IV: *ARS1* is an autonomous-replication sequence while *CEN4* provides centromere function. The *TEL* sequences are derived from the termini of the *Tetrahymena* macronuclear ribosomal DNA (rDNA) molecules. The vector was constructed as follows. The Sma I site in the *URA3* gene of YIp5 (*36*) was deleted by digestion with Ava I followed by religation (*37*); the resultant plasmid was cleaved at the single Pvu II and Eco RI sites present in its pBR322-derived sequences (*38*) and the Eco RI end was filled in with the Klenow fragment of DNA polymerase I to produce a blunt-ended 3.2-kb fragment. This fragment was ligated to a 5.5-kb fragment, containing *TRP1, ARS1*, and *CEN4*, which was produced by Pvu II cleavage of YCp19, a plasmid derived by Mann and Davis by cloning a Bam HI–Eco RI fragment from Sc4137 into the pBR322-derived sequences of YRp17 (*39*). The resultant plasmid, pPM662, contains the regenerated Pvu II site of pBR322 and its adjacent replication origin and ampicillin-resistance genes, the *TRP1-ARS1-CEN4* region of YCp19, and the *URA3* gene of YIp5. The *SUP4-o* gene was cloned into the filled-in Bam HI site of pPM662 on a 262-bp Alu I fragment derived from pSU4-A (*40*) by way of synthetic Sfi I–Not I linkers whose sequence was GCGGCCGCXGCGGCCGC (X is a mixture of G and C); the resultant plasmid was named pPM664. A short portion of the *CEN4* region of pPM664, containing an unwanted Xho I site, was deleted by digesting the plasmid at the nearby sites for Eco RI and Kpn I; the Klenow fragment of DNA polymerase I was used to create blunt ends by filling in the 5′ extension left by Eco RI and degrading the 3′ extension left by Kpn I, and then religating. This procedure regenerated an unwanted, new Eco RI site that was destroyed by cleavage with Eco RI, filling in, and religation to create pPM668. Finally, pYAC2 was created by inserting the *TEL-HIS3-TEL*-Xho I fragment of A240 p1 (provided by A. Murray) into the pBR322-derived Pvu II site of pPM668 by way of Xho I linkers; A240p1 contains the same *TEL-HIS3-TEL*-module as the plasmids A252p6 (*15*) and A142p1 (*19*).

single genetic locus, whose protein-coding function could be fulfilled by as little as 15 kb of DNA, actually covers more than a million base pairs (*3*).

Although techniques exist for cloning large genes or gene clusters in many overlapping pieces, this process is laborious, prone to error, and poorly suited to functional studies of the cloned DNA. Furthermore, there are a number of problems in molecular genetics that require the characterization of even more extensive tracts of DNA than those present in the largest known genes. For example, the regulated somatic DNA rearrangements that give rise to functional immunoglobulin genes and T cell receptor genes involve deletions of whole segments of chromosomes, while some of the genetic events that have been implicated in the induction or progression of malignant tumors involve the amplification or deletion of similarly large regions (*4–6*). In other instances, including efforts to define the primary defects in such genetic diseases as Huntington's chorea and cystic fibrosis, only the genetic linkage between the closest identified clones and the disease locus is known (*7, 8*); in typical cases, the search for the locus itself will require the analysis of megabase-pair regions of DNA. Finally, there is increasing interest in the global mapping of the DNA of intensively studied organisms (*9, 10*). Particularly in the case of the human, or other organisms with comparably complex genomes, such projects would require the ordering of hundreds of thousands of conventional clones. A cloning system that allowed the same objective to be achieved with many fewer clones would not only improve mapping efficiency but might also have dramatic effects on the reliability and continuity of the final map.

We report here the development of a high-capacity cloning system that is based on the in vitro construction of linear DNA molecules that can be transformed into yeast, where they are maintained as artificial chromosomes. Several considerations favored this combination of replicon and host. The basic functional units of yeast chromosomes—centromeres, telomeres, and ARS's (autonomous-replication sequences, with properties expected of replication origins)—have all been defined (*11–13*). In each case, DNA segments that display full functional activity in vivo are confined, at most, to a few hundred base pairs. When these elements are combined on artificial chromosomes that are approximately 50 kb or larger, the chromosomes display enough mitotic and meiotic stabilities to make their genetic manipulation straightforward (*14, 15*). Larger artificial chromosomes, consisting primarily of concatemers of bacteriophage lambda DNA, display increased mitotic stability as they increase in size, a conclusion that is reinforced by comparable studies on natural yeast chromosomes whose structures have been manipulated by homology-directed transformation experiments (*16–19*).

These prior studies of yeast artificial chromosomes have been done on molecules that formed in vivo by recombination between very small linear plasmids and transforming DNA. Typically, the linear plasmid contained all sequences required for replication and segregation, while the transforming DNA contained a marker that can be used to select for the recombinant molecule. Although such systems allow artificial chromosomes to be custom-tailored for genetic studies, they are poorly suited to cloning applications. In contrast, the system described here involves only the standard steps associated with conventional cloning protocols: in vitro ligation of vector and source-DNA fragments followed by transformation of the intact replicons into host cells.

Vector system. The vector (Fig. 1) incorporates all necessary functions into a single plasmid that can replicate in *Escherichia coli*. This plasmid, called a "yeast artificial chromosome" (YAC) vector, supplies a cloning site within a gene whose interruption is phenotypically visible (*SUP4*), an ARS (*ARS1*), a centromere (*CEN4*), selectable markers on both sides of the centromere (*TRP1* and *URA3*), and two sequences that seed telomere formation in vivo (labeled *TEL*). As described by Murray *et al.*, cleavage at the Bam HI sites adjacent to the *TEL* sequences produces termini that heal into functional telomeres in vivo (*17*).

The overall cloning protocol is shown schematically in Fig. 1. Double digestion of the particular YAC vector shown, pYAC2, with

Bam HI and Sma I yields three parts, which can be regarded as a left chromosome arm, including the centromere, a right chromosome arm, and a throwaway region that separates the two *TEL* sequences in the circular plasmid. The two arms are treated with alkaline phosphatase to prevent religation, and then ligated onto large insert molecules derived from the source DNA by partial or complete digestion with an enzyme that leaves Sma I–compatible (that is, blunt) ends. The ligation products are then transformed into yeast spheroplasts by standard methods, which involve embedding the transformed spheroplasts in agar on a selective medium.

Primary transformants are selected for complementation of a *ura3* marker in the host by the *URA3* gene on the vector. The transformants are screened for complementation of a host *trp1* marker, which ensures that the artificial chromosomes have derived both their arms from the vector, and for loss of expression of the ochre suppressor *SUP4*, which is interrupted by insertion of exogenous DNA at the Sma I cloning site, a naturally occurring restriction site in the region coding for *SUP4*'s $tRNA^{Tyr}$ gene product. *SUP4* is a particularly advantageous interruptible marker since, in an *ade2-ochre* host, cells that are expressing the suppressor form white colonies and those in which the suppressor has been inactivated form red colonies (*20*).

Pilot experiments. An initial test of the vector system involved cloning Sma I limit-digest fragments of yeast and human DNA into pYAC2. The limit-digest fragments produced by cleaving either of these DNA's with Sma I, which recognizes the sequence CCCGGG, are predominantly in the size range 20 to 200 kb. In the case of the human digest, the insert DNA was size-fractionated by velocity sedimentation to eliminate fragments smaller than 40 kb. Because we anticipated a low cloning efficiency, the ligation mixtures were carried out on a large scale (50 μg of vector plus 25 μg of insert in 200 μl).

The initial yeast-into-yeast experiment produced only a handful of transformants of the desired phenotype (Ura^+, Trp^+, Ade^-, can^R, red), but a high fraction of those obtained contained novel DNA molecules that behaved electrophoretically as though they were linear DNA molecules between 40 and 130 kb in size. Control experiments suggested that the main reason for the low efficiency was that the host strain transformed poorly, even with conventional *Escherichia coli*–yeast shuttle vectors. For the cloning of human sequences into yeast, a new host strain, AB1380, was employed with dramatically improved results. When only half the ligation mixture was transformed into 5×10^7 cells, 1×10^4 Ura^+ transformants were obtained. In a sample of 48 randomly picked colonies, 28 had all the phenotypes expected for bona fide recombinants, while 16 contained artificial chromosomes large enough to detect on ethidium bromide(EtBr)–stained pulsed-field gels (>40 kb). This sampling suggested that the experiment had an overall yield of 300 usable clones per microgram of insert DNA.

Structure of representative clones. A number of clones from the yeast-yeast and human-yeast pilot experiments were analyzed in more detail to determine whether or not the artificial chromosomes that had been produced had the expected structures (Fig. 2). The artificial chromosomes are visible on the EtBr-stained gel as 50- to 130-kb molecules, migrating ahead of the smallest natural yeast chromosome (Fig. 2A). The DNA was transferred from this gel to nitrocellulose and assayed sequentially by DNA-DNA hybridization with ^{32}P-labeled probes prepared from plasmid pBR322 (Fig. 2B) and total human genomic DNA (Fig. 2C). All five artificial chromosomes hybridized to pBR322, by way of the pBR322 sequences

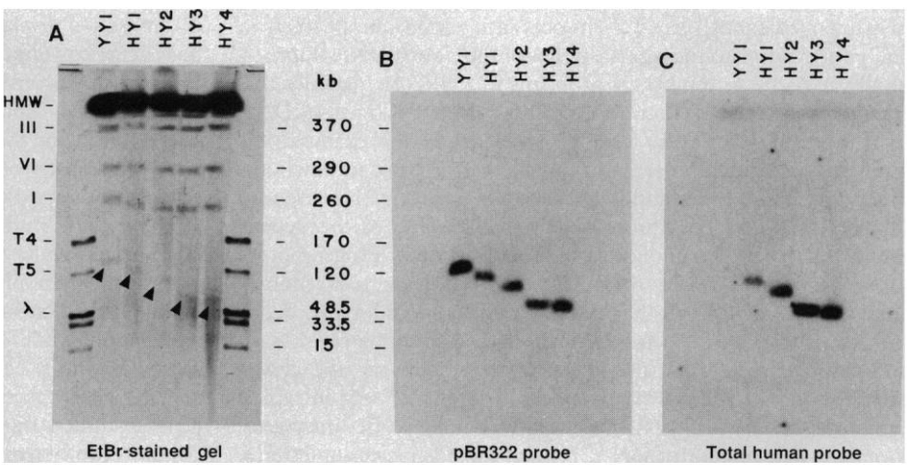

Fig. 2. Characterization of five YAC clones by pulsed-field gel electrophoresis. YY1 (yeast–yeast 1) is a clone containing yeast DNA cloned into yeast; the HY (human-yeast) clones contain human DNA cloned into yeast. (**A**) An EtBr-stained pulsed-field gel of the transformants, in which the artificial chromosomes are visible as faint bands migrating ahead of chromosome 1, which at 260 kb is the smallest natural yeast chromosome (*33, 34, 41*). The separation was made on a modified OFAGE apparatus (*34*) with a pulse time of 20 seconds; samples were prepared in agarose plugs, as described (*33, 41*). (**B**) An autoradiogram showing hybridization of all five clones to ^{32}P-labeled pBR322 by way of vector-derived sequences. For this experiment, the DNA in the gel shown in (A) was transferred to a nitrocellulose sheet as described by Southern (*42*); the probe was labeled by the hexamer-priming method (*43*). (**C**) An autoradiogram showing hybridization of only the four HY clones to ^{32}P-labeled total-human DNA. The radioactivity associated with the pBR322 hybridization was stripped off the filter before rehybridization with the total human DNA probe. All the clones were produced by ligating pYAC2-derived "arms" to source DNA that had been digested to completion with Sma I and transforming the ligation mixtures into AB1154 [used for YY1; *MATa* ψ^+ *ura3 trp1 ade2-1 can1-100 lys2-1 met4-1 tyr1; ade2-1, can1-100, lys2-1*, and *met4-1* are ochre alleles, suppressible by *SUP4-o*; ψ^+ is a cytoplasmic determinant that enhances suppression (*44, 45*)] or AB1380 (used for HY clones; *MATα* ψ^+ *ura3 trp1 ade2-1 can1-100 lys2-1 his5*). The yeast DNA, which was derived from strain AB972, was prepared as described (*34*); the human DNA, which was derived from the neuroblastoma cell line NLF, was provided by G. Brodeur. Both DNA samples had been prepared as liquid solutions; nonetheless, the average size of the fragments present before cleavage exceeded 500 kb. After Sma I digestion, the human sample was size-fractionated on a sucrose gradient. Fractions larger than 40 kb were pooled. For the Sma I digests, 25 μg of source DNA was digested to completion, gently extracted first with phenol and then with chloroform, dialyzed against TE8 (10 m*M* tris-HCl, 1 m*M* EDTA, *p*H 8), and concentrated in a collodion bag concentrator (Schleicher & Schuell UH 100/1). Vector DNA was prepared by digesting 50 μg of pYAC2 DNA to completion with Sma I and Bam HI, treating with an excess of calf-intestinal alkaline phosphatase (Boehringer Mannheim, molecular biology grade), extracting with phenol and then chloroform, and concentrating by ethanol precipitation; the throwaway Bam HI fragment containing the *HIS3* gene was not separated from the other two vector fragments. The ligation reaction was carried out for 12 hours at 15°C in a volume of 200 μl with 50 units of T4 ligase (Boehringer Mannheim) in 50 m*M* tris-HCl, 10 m*M* $MgCl_2$, 1 m*M* adenosine triphosphate, *p*H 7.5; after ligation, the reaction mixture was subjected to sequential extractions with phenol and chloroform and dialyzed against TE8. Half the ligation mixture was transformed into 5×10^7 cells, which had been converted to spheroplasts with lyticase, and plated onto four 100-mm petri plates with the use of a synthetic spheroplast-regeneration medium lacking uracil (*46*); the transformation protocol was as described (*47*).

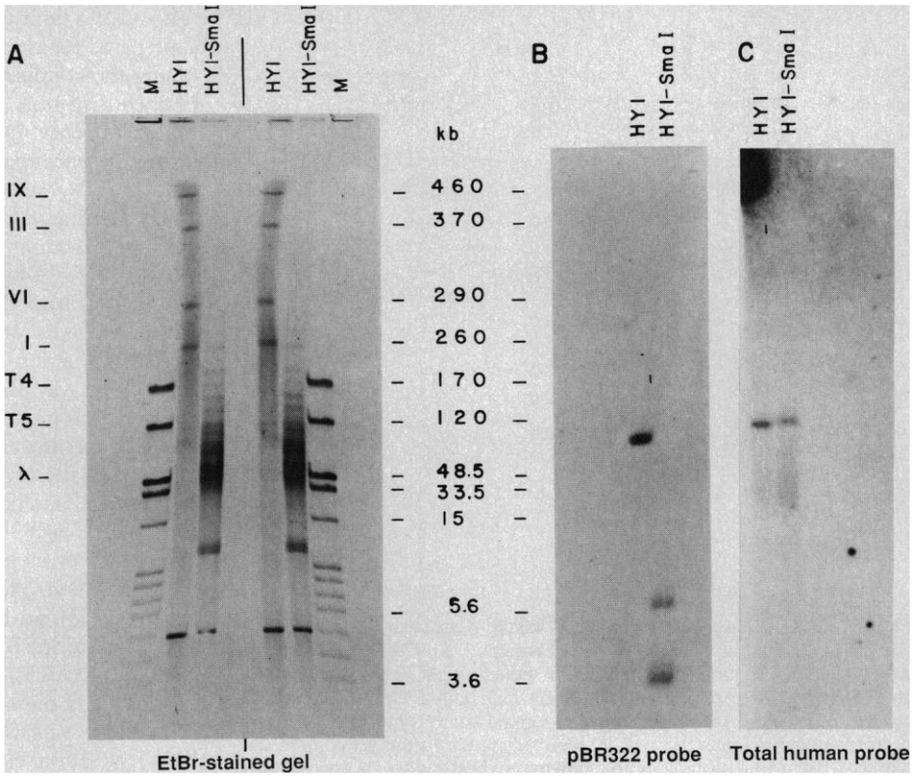

Fig. 3. The insert in the human-yeast clone HY1 is a single large Sma I fragment. (**A**) An EtBr-stained gel on which uncleaved and Sma I–cleaved DNA from the transformed yeast strain has been fractionated along with size markers (M) in two identical sets of lanes. The high background and absence of intact chromosomes larger than IX (460 kb) in the uncleaved DNA is accounted for by the use of liquid DNA samples (*34*) in this experiment. (**B**) An autoradiogram showing that, in the uncleaved sample, ^{32}P-labeled pBR322 hybridizes to intact HY1; but in the Sma I digest it hybridizes only to the 5.6- and 3.6-kb arms contributed by pYAC2. The left side of the gel shown in (A) was used in this experiment; methods were similar to those described for Fig. 2. (**C**) An autoradiogram showing that in both the uncleaved and Sma I–cleaved samples, total human DNA hybridizes to DNA molecules of similar size; in the cleaved sample, the band corresponds to the single human Sma I fragment present as an insert in HY1. The right side of the gel shown in (A) was used in this experiment.

present in pYAC2. As expected, only the four human-yeast chromosomes hybridized to human DNA; under the hybridization conditions employed, the hybridization to total human DNA is expected to involve primarily dispersed repetitive human sequences present both in the probe and in the cloned segments of human DNA. The absence of minor bands in Fig. 2, even when the autoradiograms are overexposed, suggests that the yeast artificial chromosomes are propagated faithfully.

Further analysis of two of the larger artificial chromosomes, HY1 and YY1, demonstrated that both clones have the structures expected for molecules formed by the simple pathway shown in Fig. 1. For example, the data in Fig. 3 demonstrate that the insert in HY1 can be released from vector sequences as a single 120-kb fragment by Sma I digestion. In Fig. 3A, an EtBr-stained gel is shown on which two sets of samples of uncleaved and Sma I–cleaved yeast DNA from the transformant containing HY1 have been fractionated on two identical half-gels. The DNA was transferred to nitrocellulose; the samples on the left were assayed with a pBR322 probe (Fig. 3B) and those on the right were assayed using a total-human probe (Fig. 3C). The pBR322 probe detects the intact artificial chromosome in the uncleaved sample but only the two short vector arms in the Sma I–cleaved sample. In contrast, while the total-human probe again detects the intact chromosome in the uncleaved sample, it detects a single large fragment in the Sma I–cleaved sample, which is not significantly different in size from that of the intact chromosome (*21*). Similar results were obtained for YY1.

Comparison of homologous cloned and genomic Sma I fragments. As a more stringent test of whether or not HY1 and YY1 represent authentic clones, we also showed that the Sma I fragments cloned in these artificial chromosomes are the same size as homologous fragments in the source DNA. This test required the isolation of DNA fragments from the YAC inserts that could be used to probe size-fractionated Sma I digests of genomic human and yeast DNA. We isolated these fragments by a plasmid-rescue technique that takes advantage of the presence of pBR322 sequences in the original YAC vector. Digestion of insert-containing artificial chromosomes with Xho I would be expected to produce an Xho I fragment starting adjacent to the left *TEL* sequence and extending into the insert to the first Xho I site in the cloned DNA (Fig. 1). Such a fragment would contain the pBR322 origin of replication and ampicillin resistance gene (labeled ori and Amp in Fig. 1), which are the only portions of the plasmid that are essential for replication and selection in *E. coli*. In practice, plasmid rescue is a three-step procedure: total yeast DNA from transformants is digested with Xho I, ligated under conditions that favor formation of monomer circles, and transformed into *E. coli* with selection for ampicillin resistance. Plasmids with the expected structures were readily isolated by this method from both YY1 and HY1; for example, the YY1-derived plasmid was used to demonstrate that the large Sma I fragment cloned in this artificial chromosome has a counterpart in the source DNA (Fig. 4). When DNA was transferred from the gel shown in Fig. 4A and assayed by hybridization with the *E. coli* plasmid that contained sequences rescued from YY1, the results confirmed that the Sma I fragment cloned into YY1 is of the same size in the clone and in genomic yeast DNA. In AB972, the probe hybridizes to a large chromosome in the uncleaved DNA (identified as XII in a separate experiment) and to a 120-kb fragment in the Sma I–cleaved DNA. In the sample containing uncleaved DNA from the transformed strain (lane 3), it hybridizes both to the large natural chromosome and to the small artificial chromosome, while in the sample containing cleaved DNA from the transformed strain (lane 4), it hybridizes only to the 120-kb Sma I fragment. In the latter sample, the band represents a direct superposition of homologous Sma I fragments released from the natural and artificial chromosomes. Similar experiments were carried out with a probe rescued from HY1; the only discrete fragment to which this probe hybridized in Sma I digests of human DNA was of the correct size, but the probe also weakly cross-hybridized to a heterogeneous smear of smaller Sma I fragments.

Indirect end-label mapping of YAC clones. With any primary

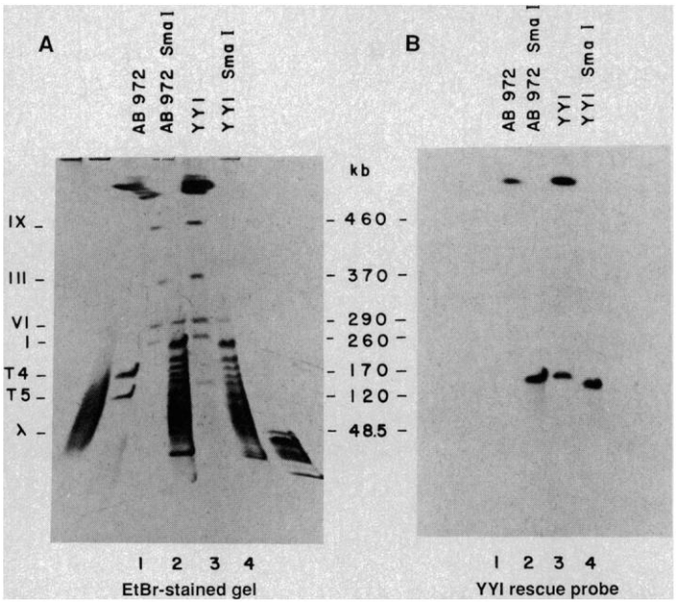

Fig. 4. The large yeast Sma I fragment cloned into the yeast-yeast clone YY1 is also present in the genome of the yeast strain AB972 from which the source DNA was extracted. (**A**) An EtBr-stained OFAGE (orthogonal field alternation gel electrophoresis) (*34*) gel on which both uncleaved and Sma I–cleaved DNA from AB972 and the transformed strain containing YY1 have been fractionated (lanes 1 to 4). The outside lanes on the gel contain size markers. (**B**) An autoradiogram showing hybridization of the samples described in (A) to a plasmid probe "rescued" from the insert of YY1. This probe hybridizes in lane 1 to sequences present at the normal chromosomal site of the cloned Sma I fragment in AB972; the large hybridizing chromosome was identified as XII in a separate experiment. In lane 2, it hybridizes to a single large Sma I fragment released from chromosome XII of AB972. In lane 3, it hybridizes both to the transformation host's chromosome XII and to the YY1 itself. In lane 4, it hybridizes to the comigrating Sma I fragments released from the transformation host's chromosome XII and from YY1.

cloning system, it is essential to have relatively simple methods of surveying the restriction sites present in newly isolated clones. Because of their large sizes, YAC clones are difficult to map by standard techniques. They are, however, particularly well suited to indirect end-label mapping, a method that has been widely employed for genomic sequencing and the mapping of hypersensitive sites in chromatin (*22, 23*). For indirect end-label mapping, a partial digest of the chromosome is size-fractionated and then probed with an end-adjacent sequence, thereby revealing a ladder of bands, the sizes of which correspond to the distances from the end to the various cleavage sites for the restriction enzyme. YAC clones are particularly well suited to indirect end-label mapping because (i) they have natural ends, (ii) the pBR322-derived sequences adjacent to each telomere allow redundant mapping of all clones with just two universal probes; and (iii) the need to detect partial-digest fragments that are present in much less than single-copy amounts is facilitated by the low sequence complexity of yeast DNA (0.5 percent of the mammalian case).

All the sites throughout a 100-kb segment could be mapped with just the left-end probe, on a single high-resolution field-inversion gel (Fig. 5). Complementary data were obtained with a right-end probe, which allowed completion of the map and also provided confirmation of the whole central region of the map. Although separate gels were used in these experiments, the two probes could equally well have been used sequentially on the same filter.

Improved vectors. The most serious limitation on pYAC2 as an all-purpose YAC vector is the inflexibility of the cloning site. Sma I produces blunt ends, which do not ligate as efficiently as "sticky" ends and also limit the range of ways in which the source DNA can be prepared for cloning. Although various methods exist for overcoming these limitations, they are all more complex than a simple sticky-end ligation and carry the attendant risk of reducing the cloning efficiency. In seeking to adapt YAC vectors to the cloning of fragments generated by a variety of restriction enzymes that leave cohesive ends, we were able to preserve the attractive features of *SUP4* as an interruptible marker—while circumventing the inflexibility of the coding-region Sma I site—by moving the cloning site into the gene's 14-bp intron. We expected that the need to maintain *SUP4* function in the vector would place few constraints on the intron's sequence or precise length, but that the cloning of huge inserts into the intron would still inactivate the gene. Site-directed mutagenesis was used to make a single nucleotide change in the wild-type intron, creating a Sna BI site that occurs only once in the vector. In the process of reconstructing a vector that contained the Sna BI site, the Sfi I and Not I sites flanking the *SUP4* gene in pYAC2 were eliminated in order to allow these enzymes to be used more conveniently either in preparing source DNA for cloning or in the analysis of clones. In this way, a new series of vectors was constructed that offer the following cloning sites: the Sna BI site itself (pYAC3), an Eco RI site created by insertion of an Eco RI linker into the Sna BI site (pYAC4), and a similarly constructed new Not I site (pYAC5). In all cases, the cloning sites occur only once in the vector and these manipulations preserved *SUP4* function.

The most extensively tested of the new vectors is pYAC4, which allows the direct cloning of inserts produced by Eco RI partial digestion. We tested this vector on a population of Eco RI partial-digest fragments prepared from the DNA of circulating human leukocytes. The uncleaved source DNA had an average size of more than 1000 kb, while the partial-digest fragments, which were not size-fractionated, were predominantly in the size range 50 to 700 kb. In this experiment, the cloning efficiency was similar to that reported above for pYAC2 cloning (several hundred clones per microgram of source DNA), but the proportion of the primary transformants that had all the phenotypes expected of bona fide recombinants was much higher (>90 percent); furthermore, nearly all such colonies contained a single artificial chromosome that hybridized to human DNA. For example, when DNA was prepared from ten clones that were picked at random all ten clones proved to contain human DNA in artificial chromosomes ranging in size up to more than 400 kb (Fig. 6).

Future prospects. The above data suggest that the generation of yeast artificial chromosomes with YAC vectors may provide a general method of cloning exogenous DNA fragments of several hundred kilobase pairs. Although only "anonymous" clones have been analyzed, these test cases appear to be propagated as faithful copies of the source DNA. The efficiency with which clones can be generated is ample to allow generation of multi-hit comprehensive libraries of the genomes of higher organisms, particularly in applications in which the availability of source DNA is not limiting. In such situations, the number of clones that can be obtained per petri plate after transformation is the most relevant measure of practicality; the present procedures with YAC vectors allow the recovery of thousands of clones per plate, a number that compares favorably with the need for 2×10^4 clones with 150-kb inserts to obtain single-hit coverage of a mammalian genome.

Further progress toward characterizing the YAC cloning system and applying it to specific biological objectives requires the development of efficient methods of screening libraries for sequences that are present in a single copy in the source DNA. Colony screening methods have been described for yeast (*24, 25*), and the screening of YAC libraries should not pose fundamental difficulties. There are, however, several practical issues that are likely to make the process

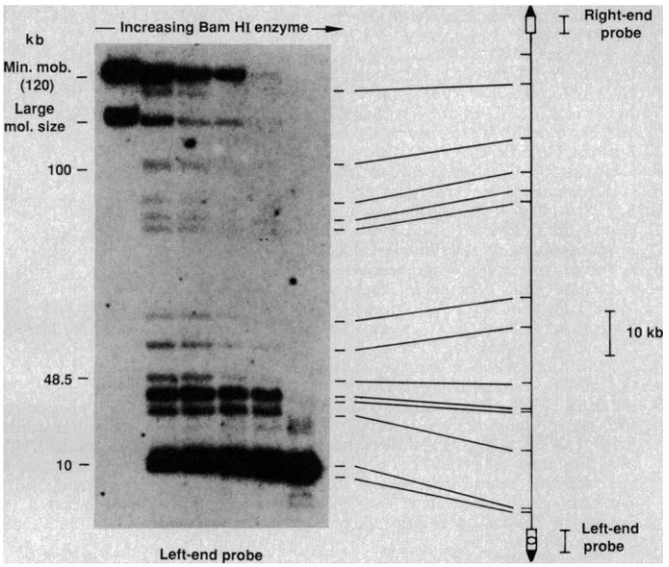

Fig. 5. Indirect end-label mapping of YY1 with Bam HI. DNA prepared from the transformant containing the artificial chromosome YY1 was subjected to partial digestion with increasing concentrations of Bam HI. The samples were fractionated by field-inversion gel electrophoresis under conditions that maximize resolution in the size range 50 to 110 kb. The DNA was transferred to nitrocellulose and assayed by hybridization with a probe consisting of vector sequences present at the left end, as defined in Fig. 1, of all YAC clones. The composite map shows the interpretation of the bands in the autoradiogram in terms of Bam HI sites in YY1. The correlation between bands and sites requires careful size calibration of the gel since gels of this type show a complex relation between size and mobility: there is a substantial compression in the range 15 to 30 kb and a short double-valued region at the top of the gel (molecules of 120 kb have minimal mobility while a heterogeneous population of larger molecules comigrate at a slightly higher mobility; there is some nonspecific hybridization to the large accumulation of yeast DNA migrating in this high molecular weight band). Sites near the right end were mapped in an experiment identical to that shown, except that a probe specific for the right end of YAC clones was employed. The DNA used in these experiments was a liquid sample prepared as previously described (*34*). The electrophoresis conditions involved a field strength of 10.5 V/cm (measured in the gel), a forward pulse time of 2 seconds, a reverse pulse time of 0.667 second, and a total running time of 12 hours; other conditions were as described (*35*). The left-end probe was the larger, and the right-end probe the smaller, of the two fragments produced by double-digestion of pBR322 with Pvu II and Bam HI; both fragments were gel-purified before labeling.

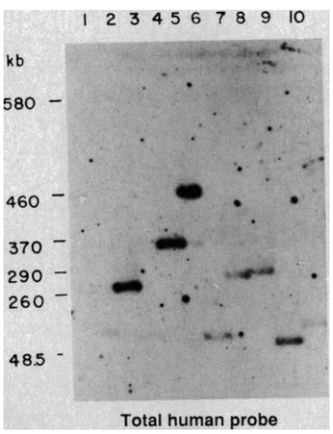

Fig. 6. Sizing of the artificial chromosomes present in ten transformants generated by cloning Eco RI partial digest fragments of human DNA into the Eco RI vector pYAC4. DNA from the transformants was prepared in agarose blocks, subjected to electrophoresis at a pulse time of 30 seconds on a pulsed-field gel apparatus that produces uniform, transverse fields intersecting at 120° (*48*), transferred to nitrocellulose, and assayed by hybridization with ^{32}P-labeled human DNA. The human DNA for the Eco RI partial digestion was prepared from circulating leukocytes by a liquid sample method whose application to yeast has been described (*34*). The average size of the DNA before partial digestion with Eco RI was approximately 1000 kb; digestion of 40 μg of this sample was carried out with 0.001 unit of Eco RI for 15 minutes at 25°C. Ligation and transformation conditions were as described for Fig. 1. The transformation host was AB1380.

more difficult than conventional colony screening (for example, both the number of the cloned molecules per cell and the number of cells per colony are much lower in YAC cloning than in plasmid or cosmid cloning in *E. coli*); the need to regenerate transformants in agar also precludes direct screening of primary transformants.

Further experience with the YAC cloning system will be required to assess such issues as the stability of clones, the extent to which the source DNA is randomly sampled, and the biological activity of the cloned DNA. Nonetheless, there are grounds for optimism that YAC vectors could even offer important advantages over standard cloning systems in these areas. In particular, there is reason to expect the yeast DNA-replication system to be more compatible with the sequence organization of typical eukaryotic DNA's than is the *E. coli* system. Essentially all sources of eukaryotic DNA that have been tested contain sequences that can function in yeast as ARS's at a spacing that is similar to that found in yeast itself (*13*). This observation suggests that the existence of a single ARS from the vector will rarely limit the amount of passenger DNA that can be accommodated. Also, the ubiquity of ARS's in eukaryotic DNA and their apparent lack in *E. coli* (*26*) may hint at a basic functional homology among the replication systems of eukaryotic organisms. Finally, although yeast has relatively little repetitive DNA compared to higher organisms, it has all the qualitative types of repeated sequence that have been described in these systems—for example, dispersed repetitive sequences (*27*), scrambled clusters of repeats (*27*), alternating purine-pyrimidine tracts (*28*), perfect palindromes (*27, 29*), long tandem arrays (*30*), and satellite-like simple sequences (*28*). Consequently, while these sequences are often difficult to clone in *E. coli* (*31*), they may pose no special problems in yeast. However, present experience suggests that few genomic sequences from higher organisms will be functionally expressed in yeast (*32*), thereby limiting the likelihood that particular sequences will be selected against because of their genetic content.

Whether or not the YAC system proves to be broadly useful, the demonstration of the basic feasibility of generating large recombinant DNA's in vitro and transforming them into easily manipulated host cells may stimulate experimentation with other combinations of replicons and hosts. There is a strong incentive to develop such systems since they are directed toward the major remaining gap in our ability to dissect the genomes of higher organisms. The resolution of both cytogenetic analysis and linkage mapping is of the order of a few megabase pairs of DNA, while cloning techniques have been limited to a size range of tens of kilobase pairs. These two levels of analysis, with their 100-fold difference in inherent resolution, are now bridged only by pulsed-field gel electrophoresis (*33–35*). The ability to isolate and amplify large DNA molecules in a simple genetic background would complement this powerful analytical technique and set the stage for expanded structural and functional studies of complex DNA sources.

REFERENCES AND NOTES

1. F. Karch *et al.*, *Cell* **43**, 81 (1985).
2. J. Gitschier *et al.*, *Nature (London)* **312**, 326 (1984).
3. A. P. Monaco *et al.*, *ibid.* **323**, 646 (1986).
4. S. Tonegawa, *ibid.* **302**, 575 (1983).
5. M. Kronenberg, G. Siu, L. E. Hood, N. Shastri, *Annu. Rev. Immunol.* **4**, 529 (1986).
6. J. M. Bishop, *Science* **235**, 305 (1987).
7. A. Beaudet *et al.*, *Am. J. Hum. Genet.* **39**, 681 (1986).
8. J. F. Gusella *et al.*, *Nature (London)* **306**, 234 (1983).
9. M. V. Olson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7826 (1986).
10. A. Coulson, J. Sulston, S. Brenner, J. Karn, *ibid.*, p. 7821.
11. E. H. Blackburn and J. W. Szostak, *Annu. Rev. Biochem.* **53**, 163 (1984).
12. J. Carbon, *Cell* **37**, 351 (1984).
13. D. H. Williamson, *Yeast* **1**, 1 (1985).
14. A. W. Murray and J. W. Szostak, *Nature (London)* **305**, 189 (1983).
15. D. S. Dawson, A. W. Murray, J. W. Szostak, *Science* **234**, 713 (1986).
16. P. Hieter, C. Mann, M. Snyder, R. W. Davis, *Cell* **40**, 381 (1985).
17. A. W. Murray, N. P. Shultes, J. W. Szostak, *ibid.* **45**, 529 (1986).

18. R. T. Surosky, C. S. Newlon, B.-K. Tye, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 414 (1986).
19. A. W. Murray and J. W. Szostak, *Mol. Cell. Biol.* **6**, 3166 (1986).
20. If necessary, it is also possible to select against nonrecombinants with the *can1-ochre* marker in the host. Cells expressing *SUP4* are canavanine-sensitive, while those lacking suppression are canavanine-resistant [R. J. Rothstein, *Genetics* **85**, 55 (1977)].
21. These data support the hypothesis that HY1 arose by the simple pathway diagrammed in Fig. 1. However, the cloning of a human insert containing internal Sma I sites would not necessarily indicate that the source DNA had been inadequately digested with Sma I or that more complex ligation events had occurred either in vitro or after transformation. Some human Sma I sites are protected from cleavage by CG methylation, and this protection would almost certainly be lost during propagation of the sites in yeast [W. R. A. Brown and A. P. Bird, *Nature (London)* **322**, 477 (1986)].
22. G. M. Church and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1991 (1984).
23. C. Wu, *Nature (London)* **286**, 854 (1980).
24. A. Hinnen, J. B. Hicks, G. R. Fink, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1929 (1978).
25. W. L. Fangman and B. Dujon, *ibid.* **81**, 7156 (1984).
26. D. T. Stinchcomb, M. Thomas, J. Kelly, E. Selker, R. W. Davis, *ibid.* 77, 4559 (1980).
27. J. R. Cameron, E. Y. Loh, R. W. Davis, *Cell* **16**, 739 (1979).
28. R. W. Walmsley, C. S. M. Chan, B.-K. Tye, T. D. Petes, *Nature (London)* **310**, 157 (1984).
29. H. L. Klein and S. K. Welch, *Nucleic Acids Res.* **8**, 4651 (1980).
30. J. H. Cramer, F. W. Farrelly, R. H. Rownd, *Mol. Gen. Genet.* **148**, 233 (1976).
31. A. R. Wyman, L. B. Wolfe, D. Botstein, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2880 (1985).
32. J. D. Beggs, J. van den Berg, A. van Ooyen, C. Weissmann, *Nature (London)* **283**, 835 (1980).
33. D. C. Schwartz *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 189 (1982); D. C. Schwartz and C. R. Cantor, *Cell* **37**, 67 (1984).
34. G. F. Carle and M. V. Olson, *Nucleic Acids Res.* **12**, 5647 (1984).
35. G. F. Carle, M. Frank, M. V. Olson, *Science* **232**, 65 (1986).
36. K. Struhl *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1035 (1979).
37. In our isolate of YIp5, the Sma I site is flanked by two closely spaced Ava I sites, only one of which is predicted by the *URA3* gene sequence [M. Rose, P. Grisafi, D. Botstein, *Gene* **29**, 113 (1984)].
38. J. G. Sutcliffe, *Cold Spring Harbor Symp. Quant. Biol.* **43**, 77 (1978).
39. C. Mann and R. W. Davis, *Mol. Cell. Biol.* **6**, 241 (1986).
40. K. J. Shaw and M. V. Olson, *ibid.* **4**, 657 (1984).
41. G. F. Carle and M. V. Olson, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3756 (1985).
42. E. M. Southern, *Methods Enzymol.* **68**, 152 (1979).
43. A. P. Feinberg and B. Vogelstein, *Anal. Biochem.* **137**, 266 (1984).
44. S. Liebman and F. Sherman, *J. Bacteriol.* **139**, 1068 (1979).
45. M. F. Tuite *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2824 (1983).
46. F. Sherman, G. Fink, C. Lawrence, *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1979).
47. P. M. J. Burgers and K. J. Percival, *Anal. Biochem.*, in press.
48. G. Chu, D. Vollrath, R. W. Davis, *Science* **234**, 1582 (1986).
49. We thank A. W. Murray and J. W. Szostak for providing helpful information on artificial chromosome construction and for the telomere-containing plasmid A240p1; R. W. Davis, for the YIp5 and YCp19 plasmids; G. Brodeur for the DNA from the human cell line NLF; D. Schlessinger for the human leukocytes; D. Garza for numerous helpful discussions; and P. Burgers for providing unpublished information and lyticase enzyme for the transformation procedure. Supported, in part, by the Monsanto Company and by NIH training grants 5T32-GM08036 and 5T32-ES07066.

10 March 1987; accepted 21 April 1987