Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 88–92, January 1994
Genetics

Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers

RAJENDRA P. KANDPAL, GEETA KANDPAL, AND SHERMAN M. WEISSMAN

Department of Genetics, Yale University School of Medicine, New Haven, CT 06510

Contributed by Sherman M. Weissman, August 30, 1993

ABSTRACT **We describe a simple and rapid method for contructing small-insert genomic libraries highly enriched for dimeric, trimeric,and tetrameric nucleotide repeat motifs. The approach involves use of DNA inserts recovered by PCR amplification of a small-insert sonicated genomic phage library or by a single-primer PCR amplification of *Mbo* I-digested and adaptor-ligated genomic DNA. The genomic DNA inserts are heat denatured and hybridized to a biotinylated oligonucleotde. The biotinylated hybrids are retained on a Vectrex-avidin matrix and eluted specifically. The eluate is PCR amplified and cloned. More than 90% of the clones in a library enriched for $(CA)_n$ microsatellites with this approach contained clones with inserts containing CA repeats. We have also used this protocol for enrichment of $(CAG)_n$ and $(AGAT)_n$ sequence repeats and for *Not* I jumping clones. We have used the enriched libraries with an adaptation of the cDNA selection method to enrich for repeat motifs encoded in yeast artificial chromosomes.**

Microsatellite repeats are highly polymorphic and have been used to construct coarse scale maps of the human chromosomes (1–5). For high-resolution genome mapping it would be valuable to have these markers at about 0.1-centrimorgan intervals, or about 30,000 markers distributed throughout the genome. Hybridization selection is an efficient method forrapidly identifying cDNA fragments complementary to sequences within large segments of genomic DNA as found in yeast artificial chromosome (YAC) inserts (6, 7). The hybridization selection approach could, in principle, be used to prepare packets of simple repeat-sequence clones representing a large fraction of those sequences present in the target YAC DNA.To do this it would be desirable to have librarires of short genomic fragments, most or all of which contained simple repeat sequencs.

Several methods have been developed to expedite the isolation of genomic fragments containing such simple repeats and some other sequence motifs (8–11). Screening genomic short fragment libraries with simple repeat oligonucleotide probes has proven successful for isolation of a large number of CA repeats (3, 12). However, this approach does not allow intensive concentration on a single region and becomes progressively less efficient as saturation is approached. Here we describe an approach that produces highly enriched CA dinucleotide repeat libraries of sufficient quality to make hybridization selection an efficient method for isolating these sequences from YACs. We also show that it can be applied to produce tri- and tetranucleotide repeat libraries and for construction of jumping libraries.

MATERIALS AND METHODS

Oligonucleotides. The following oligonucleotides were synthesized for these studies.

(*a*) CA repeat: biotin-ATAGAATAT$(CA)_{15}$,
(*b*) CA repeat: biotin-ATAGAATAT$(CA)_{22}$
(*c*) CAG repeat: biotin-ATAGAATAT$(CAG)_{15}$
(*d*) AGAT repeat: biotin-ATAGAATAT$(AGAT)_{11}$
(*e*) *Mbo* I adaptor: 5′-pGATCGCAGAATTCGCACGAGTACTAC
CGTCTTAAGCGTGCTCATGATGC-5′
(*f*) *Not* I affinity oligo: biotin-ATAGATATGTCAGATAACAGAT AACAGAT
(*g*) *Not* I adaptor: 5′-pGGCCGGTCTTATCTGTTATCTGTTATCTGTTATCTGACC
CCAGAATAGACAATAGACAATAGACAATAGACTGG–
CCGGp-5′

Construction of Small-Insert Genomic Library. Genomic DNA obtained from JY, a human B-lymphoblastoid cell line, was sonicated to obtain fragments of 0.2–1 kb. The fragments (40–50 μg) were incubated with 50 units of T4 DNA polymerase, 20 units of the large fragment of *Escherichia coli* DNA polymerase I, and 1 mM dNTPs in 300 μl at 37°C for 30 min. The reaction mixture was extracted with phenol and chloroform, precipated with ethanol, and dissolved in 10 mM Tris, pH 8.0/1 mM EDTA (TE). The DNA was treated with 50 units of *Eco*RI methylase and 80 mM *S*-adenosyl-L-methionine in 200 μl of reaction buffer for 30 min at 37°C. The methylated DNA (30 μg) was ligated to 20 μg of phosphorylated oligonucleotide containing an internal *Eco*RI site (pGCTTGAATTCAAGC). Linker ligated DNA was digested with *Eco*RI and the digest was electrophoresed to purify the DNA corresponding to fragments of 300–800 bp. This DNA (80 ng) was ligated to 1 μg of *Eco*RI-digested and dephosphorylated λgt10 vector (Stratagene). The ligated mixture was packaged and used to infect *E. coli* C600Hfl.

Amplification of Short Genomic Fragments Without Cloning. Genomic DNA from JY cells was completely digested with *Mbo* I. The digested DNA (5 μg) was ligated to an *Mbo* I adaptor oligonucleotide (15 μg) in 300 μl, with 4000 units of T4 DNA ligase. Free linkers were separated in a 1.2% SeaKem GTG agarose (FMC) gel and the ligated DNA was extracted from the gel. Ten nanograms of the linker ligated DNA was amplified in a single-primer PCR using an adaptor oligonucleotide as primer.

Affinity Capture of Di-, Tri-, and Tetranucleotide Repeat Motifs. *First-round capture.* The amplified *Mbo* I-digested genomic DNA or the PCR-amplified inserts from the small-fragment genomic phage library were denatured by heating in a boiling water bath for 10 min followed by quick chilling in an ice bath. Hybridization was carried out in 500 μl containing 10–15 μg of denatured genomic DNA and 5 μg of biotinylated oligonucleotide in 0.5 M sodium phosphate, pH 7.4/0.5% SDS at 50°C for 16–18 hr. The hybridization mixture was made up to 10 ml with 100 mM Tris, pH 7.5/150 mM NaCl (buffer A). It was then mixed with 150 mg of Vectrex-avidin (Vector Laboratories) that had been preincubated in buffer A with sonicated denatured salmon sperm DNA (100 μg/ml) for 30 min and then washed extensively with buffer A to remove unadsorbed salmon sperm DNA. After the hybridization mixture was mixed with Vectrex-avidin for 30–45

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.

min at room temperature, the supernatant was removed by centrifugation. The matrix was successively washed three times with 10 ml of buffer A at room temperature, once each with 5 ml of buffer A for 20 min at 50°C and 65°C, once with 0.1× buffer A for 20 min at 65°C, and finally with TE at 65°C for 20 min. The washes were concentrated to about 100 μl on a Centricon-30 (Amicon), and an aliquot from each concentrate was amplified by using λgt10 sequences flanking the *Eco*RI site as primers or a single oligonucleotide primer from the *Mbo* I adaptor, depending on the source of initial DNA used for hybridization. The amplified DNA was electrophoresed in a 1.2% agarose gel and blotted onto a Hybond membrane (Amersham). The filter was probed with ^{32}P-phosphorylated repeat oligonucleotide by hyubridization at 50°C (13), and the filters were exposed for autoradiography.

Second-round capture. The first-round affinity-captured material was PCR amplified and hybridized with biotinylated repeat oligonucleotide as before. The hybridization mixture was incubated with Vectrex-avidin and the specifically retained fragments were released at appropriate stringency. The wash fractions were concentrated, PCR amplified, and blotted on Hybond nylon membrane.

Construction of Small-Insert Library Enriched for Repeat Motifs. The eluate fractions showing enrichment for a particular repeat motif were PCR amplified and digested with *Eco*RI. The purified DNA (10–15 ng) was ligated to 0.5 μg of *Eco*RI-digested and dephosphorylated λgt10 arms. The ligated DNA was packaged into phage heads and used to infect C600Hfl bacteria.

PCR Amplification. All PCR amplifications were carried out in a Perkin–Elmer thermal cycler, model 9600. The reactions were carried out in 50 μl with 10 ng of template, 100 ng of each primer, 1 unit of AmpliTaq (Cetus) DNA polymerase, 2.0 mM $MgCl_2$, and 0.2 mM dNTPs. All the amplifications were carried out by mixing AmpliTaq buffer, template, primers, and AmpliTaq and incubating them in the machine at 82°C before adding dNTPs. Typical temperature cycling involved 35 cycles of 94°C, 30 sec; 55°C, 30 sec; and 72°C, 2 min. Amplification of the phage plaques was carried out by picking individual plaques, suspending them in 50 μl of 0.1 M NaCl/0.05 M Tris, pH 7.5/0.01 M $MgSO_4$, incubating them for 4–6 hr, and using 1 μl of this phage suspension as template for PCR.

Sequencing of Phage Clones. The individual phage clones were amplified as mentioned before and 1 μl of the amplified DNA was sequenced with a ^{32}P-phosphorylated primer by using a BRL double-stranded DNA cyclic sequencing kit.

Hybridization Selection of $(CA)_n$ Clones Corresponding to Specific YACs. The cDNA selection approach of Parimoo *et al.* (6) was adapted for selecting $(CA)_n$ clones specific to a particular YAC. The YACs used in this study were supplied by David Schlessinger (Washington University, St. Louis). An X chromosome-specific YAC mapping to the *F9* locus, yWXD311 (14), was separated by pulsed-field gel electrophoresis and the gel piece corresponding to DNA of 600 kb was cut out. This was incubated in buffer with *Hin*dIII, and after digestion the DNA was electroeluted. The YAC DNA was denatured and spotted onto a 2 mm × 2 mm piece of Hybond as described (6).

The Hybond filter piece was prehybridized at 65°C in 50 μl containing 5× standard saline phosphate EDTA (SSPE)/0.5% SDS and quencher DNAs [yeast DNA, *E. coli* DNA, human rDNA, human Cot-1 DNA, and poly(dI·dC)], as described (6). In addition to the above quenchers, we included $(CA)_{15}$ oligonucleotide at 0.1 μg/μl. After 20 hr of prehybridization the nylon filter was briefly washed with 2× SSPE/0.5% SDS and put in a hybridization solution containing heat-denatured PCR-amplified insert DNA from the $(CA)_n$-enriched library at 10 μg/ml and all the quencher DNAs except poly(dI·dC). Hybridization was carried out at 65°C for 30 hr. The nylon disc was washed with 2× standard saline citrate (SSC)/0.1% SDS at room temperature three times, 2× SCC/0.1% SDS at 65°C three times for 20 min, 1× SSC/0.1% SDS at 65°C once for 20 min, 0.2× SSC/0.1% SDS at 65°C once for 20 min, 0.1× SSC/0.1% SDS at 65°C twice for 20 min, and 0.1× SSC at room temperature twice. Finally, the disc was immersed in 40 μl of water in a 0.5-ml tube after removal of all the residual SSC from the disc. The tube was heated in a PCR machine at 95°C for 5 min. Twenty microliters of this eluate was used as template for amplification with phage sequences as primers. The amplified DNA (0.25 μg/ml) was subjected to a second round of selection hybridization as mentioned above. A second set of primers nested with the previous set of primers (6) was used for this amplification.

To prepare a $(CA)_n$ clone library specific to a given YAC, the selected DNA was amplified, digested with *Eco*RI, and ligated to *Eco*RI-digested λgt10 arms. A representative number of clones from this selected library were PCR amplified, radiolabeled by priming with random hexanucleotides (15), and used to probe filters containing DNA from various YACs.

Affinity Capture of *Not* I Jumping Clones. The feasibility of affinity capture for isolating *Not* I jumping clones was assessed by using known DNA samples—namely, a plasmid DNA, a cosmid DNA, and phage DNA containing a *Not* I site. The DNA from pKS plasmid (Stratagene), an anonymous cosmid clone (≈45 kb) containing two *Not* I sites, and concatamerized λgt103 containing a single *Not* I cloning site (16), were digested with *Not* I. The digested DNA (0.2 μg/ml) was circularized around a phosphorylated *Not* I adaptor (0.4 μg/ml) with T4 DNA ligase (1600 units/ml; New England Biolabs) in 5 ml of ligase buffer at 4°C for 24 hr. The ligation reaction was terminated by heating the mixture at 65°C for 20 min and concentrated in a Centricon-30 device. The DNA was digested with 20 units of *Mbo* I at 37°C for 2 hr and ligated to 3 μg of the *Mbo* I adaptor in 100 μl with 1600 units of T4 DNA ligase. An aliquot containing 5 ng of the ligated DNA was amplified in a single-primer PCR.

The amplified DNA (10 or 1 ng/ml) was heat denatured and hybridized to the biotinylated *Not* I affinity oligonucleotide (5 μg/ml) in the presence of human genomic DNA (10 μg/ml) obtained by PCR amplification of *Mbo* I-digested and adaptor-ligated human genomic DNA in 100 μl of 0.5 M sodium phosphate, pH 7.4/0.5% SDS at 65°C for 18–24 hr. The hybridization mixture was incubated with a suspension of 100 mg of Vectrex-avidin as described. The matrix was washed with 100 mM Tris, pH 7.5/150 mM NaCl three times at room temperature and once at 50°C for 20 min. The retained fragments were eluted by incubating the matrix with 10 mM Tris, pH 7.5/15 mM NaCl at 65°C for 20 min. The eluate was concentrated in a Centricon-30 and an aliquot was amplified in a single-primer PCR.

RESULTS

Construction of a Short-Fragment Genomic Library. A library corresponding to 15 × 10^6 plaque-forming units was prepared from sonicated JY DNA as described above. The library was characterized by amplifying 100 randomly chosen plaques. The insert sizes ranged between 200 and 700 bp. Overall, the library has inserts with an average size of about 500 bp and contains 2.5 genomic equivalents of inserts.

Affinity Capture for Enrichment of $(CA)_n$, $(CAG)_n$, and $(AGAT)_n$ Repeats. The affinity capture protocol is shown in Fig. 1. The conditions for affinity capture were standardized by hybridizing biotinylated $(CA)_n$, $(CAG)_n$, or $(AGAT)_n$ repeat oligonucleotides to their nonbiotinylated complementary oligonucleotides at 50°C. After hybridization, the mixture was incubated with Vectrex-avidin matrix. The matrix was then washed with 150 mM NaCl at 50°C and with 15 mM NaCl at 50°C and 65°C, successively. $(CA)_{15}$ repeat hybrids were not dissociated by 15 mM NaCl at 65°C, whereas $(CAG)_{15}$ repeat

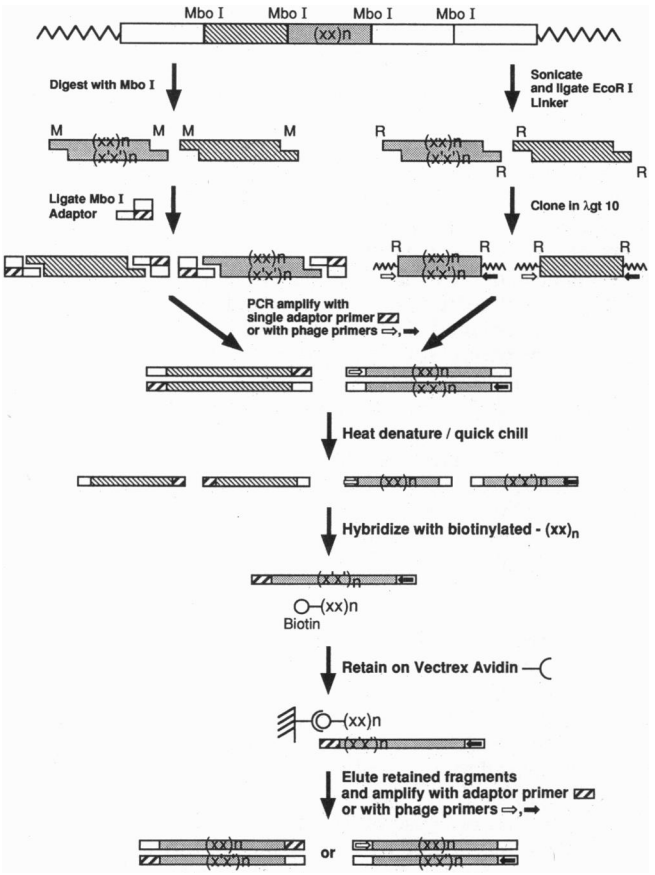

FIG. 1. Affinity-capture protocol. The first line represents genomic DNA containing several *Mbo* I sites, which are subsequently shown as M. R represents *Eco*RI overhangs obtained by addition of *Eco*RI linkers to sonicated and methylated genomic DNA, followed by digestion with *Eco*RI. The repeat motifs are designated as (XX)n.

hybrids could be eluted with 15 mM NaCl at 65°C and $(AGAT)_{11}$ repeat hybrids were dissociated in 150 mM NaCl at 65°C. Thus for $(CA)_n$ repeats, the matrix was washed with 15 mM NaCl at 65°C for 30 min, and the retained fragments were eluted with water at 65°C. The eluted DNA was amplified with an *Mbo* I adaptor oligonucleotide and processed for a second round of capture by hybridization with biotinylated $(CA)_{22}$ repeat oligonucleotide. The captured DNA was eluted with water at 65°C, amplified with *Mbo* I adaptor primer, digested with *Eco*RI, and cloned in λgt10.

For affinity capture of $(CAG)_n$ and $(AGAT)_n$ repeats, we used DNA inserts recovered from the phage DNA of sonicated JY library for hybridization with biotinylated oligonucleotides as described above. After the hybridization mixture was incubated with Vectrex-avidin, the matrix was washed with 150 mM NaCl either at 65°C or at 50°C for specific retention of $(CAG)_n$ or $(AGAT)_n$ hybrids. The affinity-captured DNA enriched for $(CAG)_n$ repeats was eluted with water at 65°C, ampified with primer set A (6), digested with *Eco*RI, and cloned in λgt10. The first-round affinity-captured DNA enriched for $(AGAT)_n$ repeats was amplified with primer set A and processed for a second round of enrichment as described in *Materials and Methods*.

Evaluation of Library Enriched for $(CA)_n$ Repeats. We randomly picked 20 plaques from the first-round enriched $(CA)_{15}$ library, amplified the insert by PCR, and blotted them onto Hybond membrane. The filter was probed with a ^{32}P-labeled $(CA)_{15}$ oligonucleotide at 50°C and washed with 0.1× SSC at 65°C for 20 min. Fifteen out of 20 clones were strongly positive for the CA repeat oligonucleotide.

A similar analysis was performed on the second-round affinity-captured DNA enriched for $(CA)_{22}$ repeat. Although hybridization of the plated library with a CA oligonucleotide gave positive signal with only 80% of the clones, PCR-amplified inserts from 16 of 17 plaques showed strong hybridization with $(CA)_{15}$ oligonucleotide (Fig. 2).

We chose 22 random clones and subjected them to DNA sequence analysis. Twenty-one clones showed runs of CA repeats, and 1 clone did not yield any sequence. The number of CA dinucleotides could be counted easily in 16 clones (Table 1). The length of CA repeat units varied between 8 and 32 with a mean value of 18.4. Nine of the randomly chosen sequenced clones were used to probe the enriched library to determine their abundance. The abundance of these plaques varied between 2 and 24, with an average of 12 (Table 1). These results indicate that there are 10,000–20,000 different $(CA)_n$ clones in our $(CA)_n$-enriched library.

Characterization of $(CAG)_n$- and $(AGAT)_n$-Enriched Affinity-Captured DNA. The $(CAG)_n$ and $(AGAT)_n$ repeats were enriched from DNA inserts recovered by PCR amplification of phage DNA of the JY sonicated library. The blots containing affinity-enriched DNA were probed with ^{32}P-phosphorylated $(CAG)_{15}$ and $(AGAT)_{11}$ oligonucleotides, respectively. The captured DNA was considerably enriched for these repeats (Fig. 2).

Twenty plaques were chosen randomly from the first-round $(CAG)_n$-enriched library for PCR amplification. Hybridization of filter blots containing amplified DNA with the labeled oligonucleotide revealed that 60% clones in the enriched library were positive for $(CAG)_n$. The first-round $(AGAT)_n$-enriched DNA was processed for a second round of enrichment, and the enriched fraction was cloned in λgt10. Six clones from the second-round enriched library were sequenced. Three of these clones contained uninterrupted runs of 4, 5, and 7 units of AGAT, respectively.

Hybridization Selection for $(CA)_n$ Specific to a Defined Region of the Genome. Ten $(CA)_n$ clones from first-round selection with YAC insert of yWXD311 were PCR amplified, blotted onto a Hybond membrane, and probed with a ^{32}P-

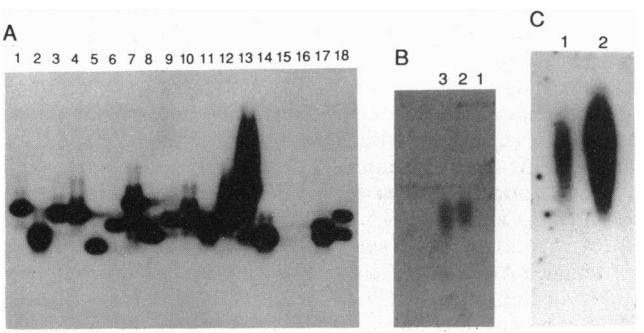

FIG. 2. Evaluation of the affinity-captured DNA. (*A*) Phage plaques from a library enriched for $(CA)_n$ repeats were used for PCR amplification. The amplified DNA was electrophoresed and blotted, and the filter was hybridized to a ^{32}P-labeled $(CA)_{15}$ oligonucleotide. The filter was washed with 0.1× SSC at 65°C for 20 min and exposed to an x-ray film. Lanes 1–18 represent clones varying in size from 200 to 800 bp. (*B*) Affinity-captured DNA enriched for $(CAG)_n$ repeats was amplified, electrophoresed, and blotted onto a Hybond filter. The filter was probed with a ^{32}P-labeled $(CAG)_{15}$ oligonucleotide, wased with 1× SSC at 50°C for 20 min, and exposed to an x-ray film. Lane 1, starting DNA material; lanes 2 and 3, affinity-enriched DNA. (*C*) Affinity-captured DNA enriched for $(AGAT)_n$ repeats was amplified, electrophoresed, and blotted onto a Hybond filter. The filter was probed with a ^{32}P-labeled $(AGAT)_{11}$ oligonucleotide, washed with 2× SSC at 50°C for 20 min, and exposed for autoradiography. Lanes 1 and 2, starting DNA and affinity-enriched DNA, respectively. Approximately equal amounts of DNA, as estimated by ethidium bromide staining, were applied to each lane.

Table 1. Characterization of clones from the CA-enriched library

Clone	Length of repeat	Abundance in 200,000 plaques
1	$(GT)_9T(GT)_{11}$	22
2	$(CA)_{18}$	24
3	$(TG)_{20}$	2
4	$(GT)_{17}$	12
5	$(CA)_{32}$	4
6	$(GT)_{22}$	10
7	$(GT)_7G(GT)_8$	14
8	$(AC)_{15}$	12
9	$(GT)_{13}$	6
10	$(TG)_{26}$	ND
11	$(CA)_{19}$	ND
12	$(CA)_{24}$	ND
13	$(CA)_{11}$	ND
14	$(CA)_{24}$	ND
15	$(CA)_{20}$	ND
16	$(GT)_{14}$	ND
Mean length*	18.4	

ND, not determined.

*An arithmetic average of the longest runs of uninterrupted stretches of CA repeats. The abundance of various clones was determined by hybridizing radiolabeled probe (15) to filters containing approximately 200,000 plaques. The repeats in the probe DNA were quenched by prehybridization (13) with $(CA)_{15}$ oligonucleotide (20 μg/ml) and human Cot-1 DNA (5 μg/ml).

labeled $(CA)_{15}$ oligonucleotide. All 10 clones hybridized to a $(CA)_{15}$ oligonucleotide. Six of these clones were sequenced. Two had identical sequence, and the remaining 4 showed different sequences. A set of primers flanking the CA repeat was synthesized for the recurring clone. These primers were used for PCR amplification of the DNA template isolatd from JY cells, target YAC, and an unrelated YAC. Although the primers appeared relatively nonspecific, as they amplified multiple bands, we observed a band of the expected size in the JY DNA and the target YAC when the PCR blot was probed with a ^{32}P-labeled CA oligonucleotide.

Twenty plaques from the second-round selected library were picked and PCR amplified, DNA was run on a gel, and the gel was blotted onto a Hybond membrane. Sixteen of the 20 plaques appeared to have CA repeats as seen by their hybridization to ^{32}P-labeled CA oligonucleotide.

We sequenced 10 clones from the second-round selected library. There were 2 clones which recurred two and three times respectively, while the remaining 5 clones were all different. The recurring clones from the second-round selection were digested with *Mnl* I and a unique half from one clone and the CA-containing half from the other clone were used to probe a DNA blot containing *Mbo* I-digested JY DNA, target YAC DNA, and an unrelated YAC DNA. There was an expected signal with the JY DNA and the target YAC DNA but no signal with the nontarget DNA (Fig. 3). The abundance of the first and second of these CA repeats in the selected library is 31% and 25%, respectively. These data demonstrate the feasibility of the selection approach for isolating a dense representation of repeat motifs specific to any chromosomal subregion isolated in YACs.

Isolation of *Not* I Jumping Clones. *Not* I jumping libraries could be constructed by digesting human genomic DNA with *Not* I and circularizing the *Not* I fragments (0.2 μg/ml) around a *Not* I adaptor oligonucleotide (0.5 μg/ml). The circularized DNA would be digested with a frequent cutter such as *Mbo* I. The *Mbo* I fragments would then be ligated to an *Mbo* I adaptor to facilitate PCR amplification of the *Mbo* I fragments. To isolate the *Not* I adaptor oligonucleotide containing *Mbo* I fragments, which would have arisen by self-circularization of large *Not* I fragments around the oligonucleotide, we would use a biotinylated *Not* I oligonucleotide for affinity capture of the *Not* I jumping clones. We have tested the affinity-capture protocol for isolating jumping clones as *Mbo* I fragments from *Not* I-digested and circularized DNA corresponding to jumps of 3 kb (pKS), 8 kb and 35 kb (*Not* I fragments of an anonymous cosmid), and 48.5 kb (self-ligated and *Not* I-digested λgt103). A band of the expected size (220 bp) was observed for pKS (Fig. 4). Similarly, two *Mbo* I fragments corresponding to cosmid jumping clones and one fragment corresponding to the λgt103 jumping clone were observed (data not shown). The selection was carried out in the presence of genomic DNA at a level sufficient to construct *Not* I jumping libraries.

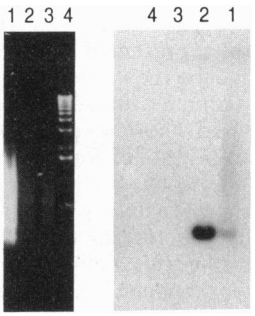

FIG. 3. Hybridization selection for CA repeat clones specific to a target DNA. A clone obtained from selection with YAC yWXD311 was digested with *Mnl* I and the two fragments were separated in a gel. The CA-containing *Mnl* I fragment was labeled with ^{32}P and hybridized to a filter blot containing *Mbo* I digest of JY DNA (lane 1), yWXD311 DNA (lane 2), and a nontarget YAC DNA (lane 3). The probe DNA was preannealed with $(CA)_{15}$ oligonucleotide (20 μg/ml) and then added to the hybridization mixture. The filter was washed with 0.1× SSC at 65°C for 30 min and exposed for autoradiography. Ethidium bromide-stained gel (*Left*) and the autoradiogram (*Right*) are shown. Lane 4 has DNA size markers.

DISCUSSION

Adaptation of the hybridization selection method for isolating YAC-specific simple sequence repeats would require an input library that does not contain significant amounts of contaminating random genomic fragments, because such fragments would tend to get selected and normalized so that the output would approximate a library of randomly cloned fragments of the target. On a random basis even a 1% contamination would result in the selection of one non-CA clone every 20–50 kb, and these clones would be expected to considerably outnumber the specific CA clones after two cycles of selection. Hybridization selection for isolating YAC-specific CA repeats may have two potential risks: (*i*) cross hybridization of CA repeats among themselves resulting in nonspecific CA clones in the selected library and (*ii*)

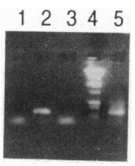

FIG. 4. Affinity capture for *Not* I jumping clones. The *Not* I digested pKS plasmid was cicularized around a *Not* I adaptor. The *Mbo* I fragment contaiing *Not* I adaptor was isolated by affinity capture. An aliquot of the affinity-captured DNA was PCR amplified and electrophoresed. Lane 1, primer alone; lane 2, 30-cycle-amplified affinity-captured DNA from the selection hybridization in a 1:1000 dilution of the pKS DNA with genomic DNA; lane 3, amplification from the hybridization containing a 1:10,000 dilution; lane 4, size markers; lane 5, 20-cycle-reamplified material of an aliquot from the sample in lane 3.

interference of *Alu* repeat elements flanking the CA repeats. However, the results presented here indicate that these factors did not influence the selection.

Ostrander *et al.* (10) have introduced an elegant approach in which a library of short single-stranded fragments is converted to a double-stranded form by priming the second-strand DNA with a CA oligonucleotide. The procedure has been reported to generate a library containing 50% CA inserts but may potentially generate more enriched material. We cannot estimate whether this approach is also able to generate libraries with a low enough background to be suitable for hybridization selection with YACs.

CA repeat sequences are useful as genetic markers because of their high polymorphism and because the incidence of polymorphisms in these markers increases with the size of the CA repeats (17). At least 80–90% of the repeats in the selected library are in the size range that is associated with a high incidence of polymorphism. The number of different inserts in the library was estimated by hybridization of single inserts to filters from the whole library. The average incidence of individual clones was 1 in 10,000–20,000. Current estimates are that the human genome may have up to 100,000 CA repeats. The lower number obtained in the present library can be attributed to two causes: (*i*) we designed conditions to select only the longer repeats and (*ii*) in the initial library, *Mbo* I restriction fragments were amplified. Because PCR amplification has a bias against recovery of larger fragments, a significant fraction of the CA repeats may have been lost.

The method described here was also successful in producing marked enrichment for simple tri- and tetranucleotide repeats, even after a single cycle of enrichment. In addition to generating further sources for polymorphic genomic markers, this approach offers an efficient alternative for identifying simple repeat sequences in cDNAs that may be subject to amplification, as has been seen in genetic syndromes associated with the phenomenon of anticipation (18).

Large-scale genome mapping has recently been accomplished most successfully by the use of macro cloning vectors, in particular YACs (14, 19–21). These approaches have tended to eclipse physical mapping methods. However, the rapid mapping methods applied to YACs have some limitations because of the occurrence of chimeric or internally deleted YACs as well as an error rate that is difficult to avoid in any large-scale procedure. Chromosome 21 is perhaps the most completely mapped autosome, and here macrorestriction mapping has provided a valuable supplement to YAC cloning procedures (22). Construction of maps by correlation of linking and jumping clones with macrorestriction fragments is a theoretically attractive way to construct physical maps (23–25), although limited by the uneven distribution of rare sites along a chromosome and the tendency of most rare cutting sites to be subject to blocking by methylation *in vivo*. In this regard the largest jumps reported in the literature are in the range of 0.5 megabase (25, 26). We have earlier described a PCR-based method for constructing linking and jumping libraries (27). The protocol presented here has the sensitivity of the earlier method and potential for greater specificity attributable to the affinity-capture step. The enrichment procedures proposed here have a potential to prepare relatively pure libraries of genomic linking and jumping clones that can be used for hybridization selection.

Our data show that affinity capture can yield libraries of clones containing inserts with any of several motifs with a relatively low background of nonspecific material. The procedures may be further extended in two ways. Use of oligonucleotide analogues that form more stable duplexes may permit isolation of groups of inserts containing short motifs of biologic importance (ref. 28; N. Baskaran and S.M.W., unpublished observation). If different libraries are prepared in vectors with divergent sequences around the cloning site, it is possible that hybridization selection can be performed simultaneously with more than one type of library. Amplification with alternative sets of primers would then produce different specific libraries of sequences complementary to a common target.

We thank Dr. David Schlessinger (Washington University) for providing the X chromosome-specific YACs used in this study. This work was supported by research funds of the American Otological Society (R.P.K.), National Institutes of Health Grant DC01682-02 (R.P.K.), and an Outstanding Investigator Award of the National Cancer Institute, CA42556 (S.M.W.).

1. Litt, M. & Luty, J. A. (1989) *Am. J. Hum. Genet.* **44**, 397–401.
2. Weber, J. A. & May, P. E. (1989) *Am. J. Hum. Genet.* **44**, 388–396.
3. Weissenbach, J., Guapay, G., Dib, C., Vignal, A., Morissette, J., Millaseau, P., Vaysseix, G. & Lathrop, M. (1992) *Nature (London)* **359**, 794–801.
4. Edwards, A., Civitello, A., Hammond, H. A. & Caskey, C. T. (1991) *Am. J. Hum. Genet.* **49**, 749–756.
5. Edwards, A., Hammond, H. A., Caskey, C. T. & Chakraborty, R. (1989) *Trans. Assoc. Am. Physicians* **102**, 185–194.
6. Parimoo, S., Pantanjali, S. R., Shukla, H., Chaplin, D. D. & Weissman, S. M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9623–9627.
7. Lovett, M., Kere, J. & Hinton, L. M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9628–9632.
8. Brenig, B. & Brem, G. (1991) *Nucleic Acids Res.* **19**, 5411.
9. Ito, T., Smith, C. L. & Cantor, C. R. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 495–498.
10. Ostrander, E. A., Jong, P. M., Rine, J. & Duyk, G. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 3419–3423.
11. Abe, K. (1992) *Mamm. Genome* **2**, 252–259.
12. Dietrich, W., Katz, H., Lincoln, S. E., Shin, H. S., Friedman, J., Dracapoli, N. C. & Lander, E. S. (1992) *Genetics* **131**, 423–447.
13. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
14. Schlessinger, D., Little, R. D., Freije, D., Abidi, F., Zucchi, I., Porta, G., Pilia, G., Nagaraja, R., Johnson, S. K., Yoon, J., Srivastava, A., Kere, J., Palmieri, G., Alfredo, C., Montanaro, V., Romano, G., Casamassimi, A. & D'Urso, M. (1991) *Genomics* **11**, 783–793.
15. Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267.
16. Lee, J. & Weissman, S. M. (1992) *Gene* **120**, 85–88.
17. Weber, J. L. (1990) *Genomics* **7**, 524–530.
18. Harper, P. S., Harley, H. G., Reardon, W. & Shaw, D. J. (1992) *Am. J. Hum. Genet.* **51**, 10–16.
19. Foote, S., Vollrath, D., Hilton, A. & Page, D. C. (1992) *Science* **258**, 60–66.
20. Bellanné-Chantelot, C., Lacroix, B., Ougen, P., Billault, A., Beaufils, S., Bertrand, S., Georges, I., Glibert, F., Gros, I., Lucotte, G., Susini, L., Codani, J., Gesnouin, P., Pook, S., Vaysseix, G., Lu-Kuo, J., Ried, T., Ward, D., Chumakov, I., Le Paslier, D., Barillot, E. & Cohen, D. (1992) *Cell* **70**, 1059–1068.
21. Chumakov, I., Rigault, P., Guillou, S., Ougen, P., Billaut, A., Guasconi, G., Gervy, P., LeGall, I., Soularus, P., Grines, L., Bougueleret, L., Bellane-Chantelot, C., Lacroix, B., Barillot, E., Gesnouin, P., Pook, S., Vaysseix, G., Frelat, G., Schmitz, A., Sambucy, J. L., Bosch, A., Estiville, X., Welssenbach, J., Vignal, A., Riethman, H., Cox, D., Patterson, D., Gardiner, K., Hattori, M., Sakaki, Y., Ichikawa, H., Ohki, M., LePaslier, D., Hellig, R., Antonarakis, S. & Cohen, D. (1992) *Nature (London)* **359**, 380–387.
22. Ichikawa, H., Shimizu, K., Saito, A., Wang, D., Oliva, R., Kobayashi, H., Kaneko, Y., Miyoshi, H., Smith, C. L., Cantor, C. R. & Ohki, M. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 23–27.
23. Smith, C. L., Lawrence, S. K., Gillespie, G. A., Cantor, C. R., Weissman, S. M. & Collins, F. S. (1987) *Methods Enzymol.* **151**, 461–469.
24. Poustaka, A. & Lehrach, H. (1986) *Trends Genet.***2**, 174–179.
25. Zabarovsky, E. R., Boldog, F., Erlandson, R., Kashuba, V. I., Ailikmets, R. L., Marcsek, Z., Kisselev, L. L., Stanbridge, E., Klien, G., Sumegi, J. & Winberg, G. (1991) *Genomics* **11**, 1030–1039.
26. Poustaka, A., Pohl, T. M., Barlow, D. P., Frischauf, A. M. & Lehrach, J. (1987) *Nature (London)* **325**, 353–356.
27. Kandpal, R. P., Shukla, H., Ward, D. C. & Weissman, S. M. (1990) *Nucleic Acids Res.* **18**, 3081.
28. Egholm, M., Buchardt, O., Nielsen, P. E. & Berg, R. H. (1992) *J. Am. Chem. Soc.* **114**, 1895–1897.