Integrative selection of human chromosome-specific yeast artificial chromosomes

(yeast artificial chromosome library/homologous recombination/genetic vector)

WILLIAM J. PAVAN* AND ROGER H. REEVES†

Developmental Genetics Laboratory, Department of Physiology, The Johns Hopkins University School of Medicine, Baltimore, MD 21205

Communicated by John W. Littlefield, June 12, 1991

ABSTRACT Human specific “integrative selection vectors” (ISVs) were designed to optimize integration of a yeast-selectable marker specifically into yeast artificial chromosomes (YACs) derived from human but not mouse DNA. ISVs were transformed into a YAC genomic library constructed from DNA of a human–mouse somatic cell hybrid containing chromosome 21 (HSA21) as the only human chromosome. One percent of the yeast in the original library contained HSA21-derived YACs; between 45% and 54% of the yeast recovered after transformation with ISV vectors contained human YACs. Integrative selection provides a rapid means of obtaining a highly enriched population of human chromosome-specific YACs by eliminating the labor-intensive steps of isolating and screening primary transformants. The procedure is biased toward the selection of YACs that contain a large number of targets for homologous recombination; thus, libraries constructed by this procedure will be composed primarily of the largest YACs in the population.

Physical analysis of mammalian genomes is simplified by using yeast artificial chromosomes (YACs) because genomic DNA inserts isolated by this approach are several hundred kilobases (kb) in length (1). YAC libraries have been generated by using DNA from human–rodent somatic cell hybrids carrying a single human chromosome, which represents 1–2% of the DNA in the cell (2, 3). This approach provides chromosome-specific human-derived YACs, but requires generation of a representative library from the hybrid (ca. 50,000 YACs with an average insert size of 300 kb for a library with 5-fold genome coverage). The 500–1000 of these YACs that have human DNA inserts are identified by picking and transferring each colony from the primary transformation reaction onto grids for two or more rounds of colony hybridization with human- and mouse-specific probes. As with other systems, yeast colony hybridization with species-specific repetitive DNA probes is subject to the isolation of both false positives (rodent YACs that are picked) and false negatives (human YACs that are not picked).

Mammalian DNA cloned as YACs can be manipulated in many of the same ways as the yeast genome. In particular, homologous recombination occurs at a very high frequency in yeast. Appropriately designed plasmids can integrate into YACs via homologous recombination with human *Alu* repetitive elements, even though they provide a small target and may diverge from a canonical *Alu* sequence by 30% (4, 5). Based on this observation, integrative selection vectors (ISVs) were designed to isolate human chromosome-specific YACs by targeting insertion of a yeast selectable marker into human but not mouse-derived YACs of a human–mouse somatic cell hybrid library. Integrative selection results in a

50-fold enrichment of human-derived YACs and eliminates several labor intensive steps in library screening.

MATERIALS AND METHODS

Yeast Strains and Plasmids. Yeast strain YPH510 carries a 360-kb human-derived YAC propagated in YPH252 (*MAT α* , *ura3-52*, *lys2-801^{amber}*, *ade2-101^{ochre}*, *trp1-1 Δ 1*, *his3- Δ 200*, *leu2- Δ 1*) (6). YWR322 carries a 350-kb mouse-derived YAC in YPH252. The YAC library used in these experiments was provided by M. K. McCormick (3). It was constructed in yeast strain YPH274 (a diploid strain homozygous for the same markers that are in YPH252) from WAV17 DNA. WAV17 is a human–mouse somatic cell hybrid line containing human chromosome 21 (HSA21) as the only human chromosome (7). Yeast centromere-containing plasmid pRS313 (8) has an autonomously replicating sequence plus the yeast selectable marker *HIS3* and can confer histidine prototrophy on yeast transformants without integration into the YAC.

ISV Construction. Cohesive ends of the 1.2-kb *EcoRI*–*BamHI* restriction fragment containing the *Saccharomyces cerevisiae HIS3* gene (9) were filled in with the Klenow fragment of DNA polymerase I and inserted into the similarly-blunted *EcoRI* sites of the Bluescript plasmid (Stratagene cloning systems). A 1.2-kb blunt-ended restriction fragment containing *HIS3* was isolated from this plasmid by using *Sma* I and *EcoRV* and ligated into the *Alu* sequence of the BLUR8 plasmid (5) at the blunt-ended *Msc* I or *Pf*MI restriction sites to generate pBP96 and pBP97, respectively (Fig. 1). pBP90 was constructed by replacing a 2.2-kb *Xba* I restriction fragment from the human long interspersed repetitive element (LINE) in plasmid pL1.1a (B. Dombrowski, S. Mathias, A. Scott, and H. Kazazian, personal communication) with the 1.2-kb *HIS3* gene by blunt-end ligation (Fig. 1). The *Xba* I fragment that was replaced to generate pBP90 corresponds to nucleotides 3392–5607 of the 6.1-kb human LINE consensus sequence (10).

Yeast Transformation and Analysis. Ten micrograms of pBP90 and pBP96 or 20 μ g of pBP97 was linearized with *Bam*HI (pBP96 and pBP97) or *Not* I/*Sal* I (pBP90), heat-inactivated at 65°C for 15 min, and transformed into 5×10^7 yeast cells made competent by the lithium acetate procedure (11). *His*⁺ transformants were selected on synthetic dextrose (SD) plates supplemented with uracil, tryptophan, leucine, lysine, and adenine but lacking histidine (SD – *His* plates). *His*⁺ transformants containing YAC color markers were

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: ISV, integrative selection vector; YAC, yeast artificial chromosome; HSA21, human chromosome 21; LINE, long interspersed repetitive element.

*Present address: Department of Molecular Biology, Princeton University, Princeton, NJ 08544.

†To whom reprint requests should be addressed at: Department of Physiology, P202, The Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, MD 21205.

FIG. 1. ISVs pBP90, pBP96, and pBP97. The yeast selectable marker *HIS3* was inserted into human LINE and *Alu* repeat sequences. The 5' and 3' symbols designate the orientation of the repetitive element relative to the insertion site of the *HIS3* gene. To reduce integration of the ISVs into the ampicillin-resistance gene (*amp^r*) and origin of replication located on YAC vector arms, the plasmid sequences are cleaved prior to transformation of yeast cells by restriction endonuclease digestion with *Bam*HI (pBP96 and pBP97) or *Not*I/*Sal*I (pBP90).

transferred from SD – His plates onto nitrocellulose circles laid on yeast/peptone/dextrose plates and processed for colony hybridization (12). Yeast DNA for repeat sequence profile analysis was prepared as described (4, 6).

Radiolabeled Probes and Hybridization. Random primer extension (13) was used to prepare probes from a 0.3-kb *Bam*HI fragment of the human *Alu* element in BLUR8 (5), a 6-kb *Not*I–*Sal*I fragment of the human LINE in pL1.1a (B. Dombrowski, S. Mathias, A. Scott, and H. Kazazian, personal communication), and a 1-kb *Pst*I fragment of the mouse LINE in B59 (14). Southern blot analysis was accomplished as described (15).

RESULTS

ISV Construction. ISVs were designed to target integration specifically into human-derived YACs by homologous recombination between the human repetitive elements *Alu* or LINE on the vectors and homologous sequences on YAC inserts (Fig. 1). *Alu*-ISVs were engineered so that the shortest possible segment is introduced into yeast in an effort to minimize deletions that could occur from targeting of adjacent *Alu* sequences. The *HIS3* gene was inserted at either of two unique restriction sites in the human *Alu* element in BLUR8 (5) (pBP96 and pBP97, Fig. 1) to make different regions of the *Alu* sequence available for homologous recombination with *Alu* elements in the YAC insert. LINES are present in both species and contain conserved and divergent regions. To reduce targeting to murine-derived YACs, a highly conserved 2.2-kb region was deleted from a human LINE and replaced with the yeast *HIS3* gene.

Human DNA-Specific Targeting. The ability of ISVs to target the insertion specifically into human DNA was assessed by transforming three artificial mixtures containing different proportions of yeast propagating human- or mouse-

derived YACs (Table 1). Independently grown cultures of yeast strains YPH510 carrying a human-derived YAC and YWR322 carrying a mouse-derived YAC were counted and combined to achieve approximately 5%, 2.5%, and 1% human YACs. The percentage of "transformable" human-derived YACs was determined by transforming an aliquot of each mixture with an autonomously replicating plasmid, pRS313, complementing the *his3* mutation in the parental yeast strain. His⁺ transformants were processed by colony hybridization with the human- and mouse-specific repeat element probes on duplicate filters. Remaining aliquots of each mixture were transformed with each of the ISV vectors independently. Each transformation was done with 5, 10, and 20 μ g of each ISV. The results obtained with the concentration providing optimal selection (10 μ g of pBP90 and pBP96; 20 μ g of pBP97) are shown in Table 1, and these amounts were used in subsequent experiments.

Transformation with the episomal control plasmid pRS313 generated a similar number of His⁺ transformants from each mixture, but ISV-transformed sets resulted in preferential recovery of cells containing the human-derived YAC (Table 1). In all nine reactions where mixtures of yeast were transformed with ISVs, insertion of the selectable marker was preferentially targeted to cells containing a human YAC and not to the more abundant yeast cells containing mouse-derived YACs. Even when YPH510 represented only 0.8% of the yeast in the transformation reaction, 36–80% of the resulting His⁺ transformants contained the human-derived YAC—a 45- to 100-fold enrichment using integrative selection.

Integrative Selection of Chromosome-Specific YACs. The feasibility of obtaining chromosome-specific YACs by means of integrative selection was assessed by transforming a portion of a HSA21–mouse somatic cell hybrid library (Table 2). Three hundred clones were chosen at random from an

Table 1. Integrative selection of human-derived YACs

Plasmid	His ⁺ transformants, no.		Human-derived YACs, %		
	YPH510 (human YAC)	YWR322 (mouse YAC)	Mix A	Mix B	Mix C
pRS313 (episomal)	1200	800	5.5 (1800)	3.8 (1200)	0.8 (1800)
pBP96 (<i>Alu</i> -ISV)	90	1	95.0 (20)	65.7 (35)	55.5 (9) [†]
pBP97 (<i>Alu</i> -ISV)	150	2	92.8 (14)	83.3 (12)	80.0 (5) [†]
pBP90 (LINE-ISV)	120	0	84.2 (19)	92.8 (14)	35.7 (28) [†]

Transformants were identified as containing a human or mouse YAC by colony hybridization. pBP96 and pBP97 transformants containing human-derived YACs were human *Alu*⁺ human LINE⁺ murine LINE[−] and those containing murine-derived YACs were human *Alu*[−] human LINE[−] murine LINE⁺. pBP90 transformants containing human-derived YACs were human *Alu*⁺ human LINE⁺ murine LINE[−] and those containing murine-derived YACs were human *Alu*[−] human LINE⁺ murine LINE⁺.

*Numbers in parentheses equal the total number of His⁺ transformants.

[†]Combined results of two transformation reactions.

Table 2. Enrichment of large insert, HSA21-derived YACs using integrative selection

Selection procedure	YAC strains analyzed, no.	Human YACs selected, no.	Human YACs in library, %	YAC sizes, kb
Colony hybridization	300	3	1	750 150 100
Integrative selection				
With pBP96 (<i>Alu</i> -ISV)	35*	18	54	72% 750
With pBP97 (<i>Alu</i> -ISV)	40*	18	45	67% 750
With pBP90 (LINE-ISV)	23*	12	52	92% 750

Transformants containing human YACs were identified by using human *Alu* profile analysis (see Fig. 2).

*Combined results of three transformation reactions.

existing library (3) and screened by colony hybridization with human repeat elements *Alu* and LINE. Five colonies hybridized to *Alu* or LINE probes in a primary screen, and three of these were confirmed as containing human-derived inserts by using repeat sequence profile analysis and pulsed-field gel electrophoresis. This is consistent with our previous results using colony hybridization to identify species-specific YACs from a somatic cell hybrid library [i.e., 40–60% of the initial isolates were false positives (unpublished observations)]. The three HSA21-derived YACs were 100, 150, and 750 kb. As expected, the two larger YACs contained proportionately more *Alu* sequences.

The 300 randomly picked clones were pooled, grown overnight in mass culture, and transformed with pBP90, pBP96, or pBP97. All His⁺ transformants obtained from each experiment were expanded for analysis; between 45% and 54% of these contained one of the human-derived YACs

(Table 2 and Fig. 2). Since the original mix of 300 YACs contained 3 with human-derived inserts, this represents a 50-fold enrichment over the original population. Different portions of the *Alu* element were exposed for homologous recombination in pBP96 and pBP97, but this did not effect selection frequency, which was similar for all three vectors.

Large YACs Are Preferentially Selected. The 750-kb YAC was disproportionately represented among the transformants with each of the ISVs, suggesting that YACs containing more repeat sequences may be preferentially targeted by integrative selection. To test this hypothesis, equal numbers of yeast carrying human-derived YACs with different numbers of *Alu* and LINE sequences were pooled and transformed with pRS313 (control) or one of the ISVs (Fig. 3). The same frequency of transformation was observed in all strains with the control plasmid, demonstrating that there was no differential effect of the YACs on transformation. The frequency of integrative selection with pBP90 increased directly with the number of LINE sequences in the YAC. Three human-derived YACs that did not contain any LINE sequences were never selected from the pool when pBP90 was used as the ISV. The same general trend was seen when *Alu* sites were targeted with pBP96, but the increased frequency was not strictly linear. A 200-kb YAC containing about 35 *Alu* sequences was selected twice as frequently as a 750-kb YAC

FIG. 2. Repeat sequence profile analysis of yeast strains isolated by integrative selection. DNA from independent His⁺ transformants obtained after integrative selection of somatic cell hybrid-derived YACs was digested with *Eco*RI and analyzed by Southern blot analysis with a human *Alu* probe. A 750-kb human YAC (lanes a, c, e, f, g, h, and l) was preferentially selected over a smaller YAC (lanes k and p). Deletions created by *Alu* ISVs were identified by loss of *Alu*-containing restriction fragments (lane f) and a decrease in size in pulsed-field gel electrophoresis analysis (not shown). DNA from yeast strains containing mouse-derived YACs (lanes b, d, i, j, m, n, and o) showed strong hybridization to the BLUR8 *Alu* sequence on the integrated ISV and weak hybridization with partially conserved murine *Alu* sequences. Positions of molecular weight markers are shown in kilobases.

FIG. 3. Effect of the number of target sites on the frequency of integrative selection. Six yeast strains containing human-derived YACs with different numbers of *Alu* sequences, ranging in size from 100 kb to 750 kb, were mixed together and transformed with a control plasmid or ISVs. His⁺ transformants were identified by repeat sequence profile analysis with human *Alu* and LINE probes on conventional Southern blots.

with 50 *Alu* sequences. The results obtained with *Alu*-targeting ISVs might be due to the occurrence on the smaller YAC of repeats that show more homology with the BLUR8 *Alu* sequence of the ISVs than did the larger YAC, providing preferred targets for homologous recombination.

Of 64 colonies containing human-derived YACs isolated with the LINE ISV (from experiments in Table 2 and Fig. 3), none had any rearrangements detectable by alterations in the pattern of *Alu*- or LINE-containing restriction fragments on the repeat sequence profile analysis. However, 16 of the 88 (18%) human-derived YACs recovered after transformation with *Alu* ISVs demonstrated loss of *Alu*-containing restriction fragments (Fig. 2, lane f) and displayed an altered mobility on pulsed-field gel electrophoresis. These deleted YACs were reduced in size by 50–400 kb and probably occurred when the ends of the ISV targeted two different *Alu* sequences, creating internal deletions.

DISCUSSION

A 50-fold enrichment of human chromosome-specific YACs from a reduced somatic cell hybrid YAC library was obtained by using ISVs to target a yeast selectable marker to the human repetitive elements *Alu* or LINE. Using colony hybridization, a comprehensive five-hit YAC library of a reduced somatic cell hybrid requires the manipulation of about 50,000 YAC-containing yeast colonies that must be picked and replated for two rounds of colony hybridization to recover the 500 (1%) human chromosome-specific YACs. The use of integrative selection eliminates the manipulation of most of the rodent-derived YACs, thereby reducing the effort required to produce a chromosome-specific YAC library. In practice, top agar from primary transformation plates can be removed and put directly into liquid culture to form pools of 500–1000 clones. Glycerol stocks of each pool provide a source of unaltered clones, which might be desirable for certain types of biological analysis. Aliquots of the pools are transformed *en masse* with ISVs and selected. Based on results presented here, 20–30 His⁺ transformants will typically be obtained, half of which will contain copies of the 5–10 human-derived YACs in the initial pool. These colonies grow on the surface of agar plates and can be directly replica-plated for a secondary screen by colony hybridization. This procedure reduces the total number of clones that must be manipulated individually by 50-fold.

The isolation of yeast that did not contain HSA21-derived inserts from the somatic cell hybrid library may have been due to integration of the ISVs into mouse-derived YACs. The human *Alu* clone, BLUR8, showed weak hybridization with many mouse-derived YACs in the repeat sequence profile analysis (Fig. 2). Murine repeat sequences that are homologous to BLUR8 may serve as low-efficiency targets for integration. The LINE ISV pBP90 was deleted for the segment that is most highly conserved between human and mouse, but regions of significant homology remain. This may be sufficient to target integration to mouse LINE sequences with low efficiency.

Alu-targeting ISVs produced deletions in 18% of the selected YACs. These events probably occur when the *Alu* sequences in the vector recombine with two different *Alu* sequences in the YAC. Since each YAC appears several times in the pool of His⁺ transformants, deleted YACs are readily detected by the loss of *Alu*-containing bands on the repeat sequence profile and can be excluded (Fig. 2). Human

YACs isolated by using the LINE-targeting ISV never contained internal deletions.

Larger YACs will include on average more targets for ISVs and have an increased probability of containing a preferred target (i.e., one that is highly homologous with the repetitive sequence present in the ISV). As a consequence, integrative selection will preferentially target large YACs. For libraries constructed with partially digested genomic DNA, every segment has an equal chance of being included on a large cloned restriction fragment, so integrative selection should create a representative collection of large YACs. The use of the ISV strategy to produce libraries from somatic cell hybrids should also reduce the frequency of chimeric YACs containing noncontiguous human sequences in the same clone. Introduction of multiple human segments into the same cell and coligation of human segments will be rare because most segments will derive from the mouse. YACs that result from the coligation or fusion of mouse and human segments will generally contain fewer targetable repeats than contiguous human segments of the same size and will be selected less frequently. Thus, cocloning events should be rare among the selected YACs. Integrative selection can be used to generate specialized YAC libraries by laboratories that lack the time or space needed to screen grided libraries because 50-fold fewer yeast transformants need to be analyzed. This approach provides a means of rapidly isolating large chromosome-specific DNA fragments useful for mapping and analysis of the human genome.

We thank Phil Hieter for suggestions and careful review of the manuscript and MaryKay McCormick for the hybrid YAC library. Cheri Byars provided excellent technical assistance. Steven Mathias and Allen Scott provided the pL1.1a plasmid; Thomas Fanning, the B59 plasmid; and Prescott Deininger, the BLUR8 plasmid. This work was supported by U.S. Public Health Service Grant HD24605-02 to R.H.R.

1. Burke, D. T., Carle, G. F. & Olson, M. V. (1987) *Science* **236**, 806–812.
2. Wada, M., Little, R. D., Abidi, F., Porta, G., Labella, T., Cooper, T., Della Valle, G., D'Urso, M. & Schlessinger, D. (1990) *Am. J. Hum. Genet.* **46**, 95–106.
3. McCormick, M. K., Shero, J. H., Cheung, M. C., Kan, Y. W., Hieter, P. A. & Antonarakis, S. E. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9991–9995.
4. Pavan, W. J., Hieter, P. & Reeves, R. H. (1990) *Mol. Cell. Biol.* **10**, 4163–4169.
5. Deininger, P. L., Jolly, D. J., Rubin, C. M., Friedmann, T. & Schmid, C. W. (1981) *J. Mol. Biol.* **151**, 17–33.
6. Pavan, W. J., Hieter, P. & Reeves, R. H. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1300–1304.
7. Kozak, C. A., Lawrence, J. B. & Ruddle, F. H. (1977) *Exp. Cell Res.* **105**, 109–117.
8. Sikorski, R. S. & Hieter, P. (1989) *Genetics* **122**, 19–27.
9. Struhl, K. (1985) *Nucleic Acids Res.* **13**, 8587–8601.
10. Scott, A. F., Schmeckpeper, B., Abdelrazik, M., Comey, C., O'Hara, B., Rossiter, J., Cooley, T., Heath, P., Smith, K. & Margolet, L. (1987) *Genomics* **1**, 113–125.
11. Itoh, H., Fukuda, Y., Murata, K. & Kimura, A. (1983) *J. Bacteriol.* **153**, 163–168.
12. Sherman, F., Fink, G. & Hicks, J. (1986) *Methods in Yeast Genetics* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
13. Feinberg, A. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267.
14. Fanning, T. G. (1983) *Nucleic Acids Res.* **11**, 5073–5091.
15. Reeves, R. H., Crowley, M. R., Lorenzon, N., Pavan, W. J., Smeys, R. J. & Goldowitz, D. (1989) *Genomics* **5**, 522–526.