Nucleic Acids Research, 1993, Vol. 21, No. 16 3911–3912

Construction of random small-insert genomic libraries highly enriched for simple sequence repeats

Luchezar Karagyozov+, Iveta D.Kalcheva and Verne M.Chapman*
Department of Molecular and Cellular Biology, Roswell Park Cancer Institute, Buffalo NY 14263, USA

Received May 24, 1993; Revised and Accepted June 29, 1993

Dinucleotide $(CA)_n$ repeat sequences are highly abundant and interspersed in eukaryotic genomes. Individual sites or loci can be identified by PCR-based assays using unique sequence oligonucleotides that flank specific CA-repeats. The number of CA-repeats at a given locus is variable making these markers highly informative for genetic analysis in humans (1) and other species (2). Unique sequences flanking specific $(CA)_n$ loci are usually identified by analyzing genomic libraries containing small size inserts, suitable for sequencing, generated by restriction enzymes. However the construction and screening of these type of libraries has some notable limitations. First, by using one or even several restriction enzymes only a subset of all $(CA)_n$ repeats can be cloned as this depends on the presence of specific restriction sites near the repeats. Second, the screening is time consuming because of the low frequency of the clones containing $(CA)_n$ repeats (~1%). To overcome the latter problem an elaborate method for library enrichment has been proposed based on selective second-strand DNA synthesis (3).

In this paper we present an efficient method for construction of random small-insert genomic libraries enriched for $(CA)_n$ repeats. The method involves fragmentation of DNA by sonication and ligation of an adaptor to the fragments followed by PCR amplification and enrichment for $(CA)_n$ repeats by hybridization to $(GT)_{15}$ Oligonucleotides bound to a nylon membrane.

Mouse DNA (10 μg) was sonicated in 2×SSC to fragment sizes of 200–600 bp. The 3'- and 5'-end phosphate groups of the fragments were removed with CIP at 53°C (4). The enzyme was inactivated and DNA precipitated. The fragments were blunt ended by T4 DNA polymerase and their 5'-ends were phosphorylated by T4 polynucleotide kinase (4). The fragments were ligated (15°C, 16 hrs) to 5'-end phosphorylated adaptor (21-mer: CTCTTGCTTGAATTCGGACTA and 24-mer: pTA-GTCCGAATTCAAGCAAGAGCACA) which carries an EcoRI site (5). Fifty ng of the ligated DNA fragments were amplified in 100 μl of 10 mM Tris–HCl, pH 8.3, $MgCl_2$ 1.5 mM, KCl 50 mM, gelatine 100 μg/ml, 200 μM dNTPs, 3 μM primer (the 21-mer) (95°C, 1 min, 55°C, 1 min, 72°C, 2 min, 30 cycles).

The fragments containing $(CA)_n$ repeats were selected by hybridization to $(GT)_{15}$ oligonucleotide bound to a nylon membrane [the immobilization of $(GT)_{15}$ is much more efficient than that of $(CA)_{15}$]. An aliquot of 1.5 μg of $(GT)_{15}$ dissolved in H_2O) was spotted onto a small (1.2–1.4 mm) square of Hybond N+, Amersham. The membrane was air dried, washed in water, baked for 2 hrs at 80°C and UV-cross-linked for 2.5 min. Unbound oligonucleotides were washed by incubation for 2 days in hybridization solution (50% formamide, 5×SSC, 50 mM Na-phosphate buffer, pH 7.0, 7% SDS) at 37°C and in 1% SDS for 10 min in a boiling water bath. The membrane was hybridized with 20 μg of denatured PCR amplified fragments in 600 μl of hybridization buffer for 48 hrs at 37°C in the presence of 10 μg of the 21-mer. The oligonucleotide excess prevents concatamerization of the fragments as their ends are complementary. After hybridization the filter was washed with 2×SSC, 1% SDS, 50 mM Na-phosphate buffer, 1×SSC and 0.1×SSC (final wash 65°C, 30 min). The bound fragments were eluted in 0.5 ml 1% SDS for 2–3 min in a boiling water bath. Glycogen (20 μg) and 10 M LiCl (100 μl) were added, the fragments were precipitated with *i*-propanol (–20°C, overnight) and dissolved in 25 μl H_2O. One μl of this preparation was amplified. The PCR product (5 μg) was used in a second 24 hrs hybridization. The degree of enrichment for fragments containing repeats was estimated by dotting the PCR products and probing with ^{32}P end-labeled $(GT)_{15}$. The results from three independent experiments showed 10–15 fold enrichment for $(CA)_n$ repeats after the first round and a further 3–5 fold enrichment after the second round of hybridization.

The degree of enrichment was also tested by cloning. To this end the amplified DNA was digested with EcoRI and electrophoresed on a 2% agarose gel. The 250 to 450 bp fragments were purified and cloned into dephosphorylated pBluescript SK+ (Stratagene). The clones were transferred to nitrocellulose and probed with ^{32}P end-labeled $(CA)_{15}$ oligonucleotide. A high enrichment of $(CA)_n$-containing clones was observed (Figure 1). After the second round of hybridization over 40% of the clones gave a positive signal.

Twenty randomly selected positive clones were sequenced. All clones contained different CA repeats positioned at least 30–40 nt from the EcoRI site. Thirteen of the clones contained perfect repeats [the shortest one was $(CA)_{10}$]. Five of the repeats were imperfect and two were compound repeats. The proportions of sequences of the various types are similar to those reported for human DNA (6), the perfect repeats predominating.

The method for library construction and enrichment which we present here compares favorably to the methods employed up to now. The mechanical fragmentation of the starting DNA makes

* To whom correspondence should be addressed

+On leave from the Institute for Cell Biology and Morphology, Bulgarian Academy of Sciences

the cloning independent of the presence of restriction sites close to the repeat. The enrichment achieved is about 50 fold. This makes the search for new CA repeat length polymorphisms much more efficient. The method described can also be applied for isolation of clones containing other simple repeated sequences present in the genome (7).

ACKNOWLEDGEMENT

This work was supported by USPHS grant HGO0170 to V.M.C.

REFERENCES

1. Hudson,T.J. *et al.* (1992) *Genomics* **13**, 622–629.
2. Dietrich,W. *et al.* (1992) *Genetics* **131**, 423–447.
3. Ostrander,E.A., Jong,P.M., Rine,J. and Duyk,G. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 3419–3423.
4. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring, Harbor, NY.
5. Duguid,J.R. and Dinauer,M.C. (1990) *Nucleic Acids Res.* **18**, 2789–2792.
6. Weber,J.L. (1990) *Genomics* **7**, 524–530.
7. Edwards,A. *et al.* (1992) *Genomics* **12**, 241–255.

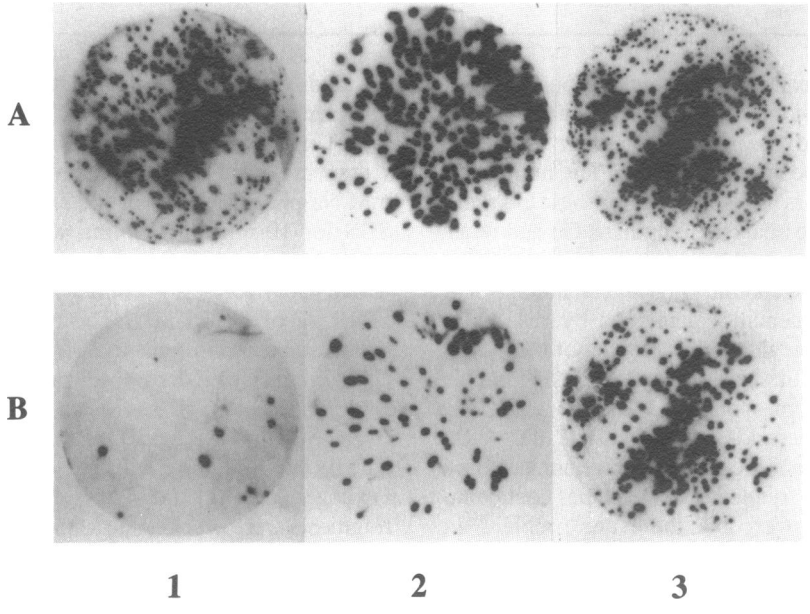

Figure 1. Enrichment of the random small-insert libraries for CA-repeats. Genomic DNA fragments generated by sonication were cloned into pBluescript prior to enrichment (*1*), after one round (*2*) and after a second round (*3*) of enrichment. The filter lifts were probed with radiolabeled vector sequences (*A*), stripped and then probed with $(CA)_{15}$ (*B*).