Use of repetitive DNA probes as physical mapping strategy in *Caenorhabditis elegans*

Giuseppina Cangiano, Humaira Ameer¹, Robert Waterston¹ and Adriana La Volpe

International Institute of Genetics and Biophysics, CNR, Naples, Italy and ¹Department of Genetics, Washington University School of Medicine, St Louis, MO, USA

Received June 11, 1990; Revised and Accepted August 8, 1990

ABSTRACT

A method for linking genomic sequences cloned in yeast artificial chromosomes (YACs) has been tested using *Caenorhabditis elegans* as a model system. Yeast clones carrying YACs with repeated sequences were selected from a *C. elegans* genomic library, total DNA was digested with restriction enzymes, transferred to nylon membranes and probed with a variety of repetitive DNA probes. YAC clones that overlap share common bands with one or more repetitive DNA probes. In 159 YAC clones tested with one restriction enzyme and six probes 28 overlapping clones were detected. The advantages and limitations of this method for construction of YAC physical maps is discussed.

INTRODUCTION

The cloning of large DNA fragments into yeast artificial chromosomes (YACs) has several advantages for the construction of physical maps (1). These include the large size of cloned fragments (100–1000 kb or more) especially with accurate pre-sizing (3) and the possibility, in combination with other kinds of libraries (1), of increasing the representation of the genome; it is likely that sequences difficult to clone or maintain in a prokaryotic host may be stable in a eukaryotic chromosome. The utility of YACs in generating long range physical maps has been demonstrated in the *C. elegans* mapping project. The *C. elegans* physical map was initially constructed using a fingerprinting method to detect overlaps between cosmid clones and to obtain groupings of segments of DNA in contigs. The analysis of 17500 cosmids (6.5 genomic equivalents) yielded a map with more than 700 contigs (1) with an average length of about 125kb. This approach reached its practical limit when 90–95% of the genome had been cloned. To fill the gaps between contigs it has been necessary to use a library constructed in yeast artificial chromosomes (YACs) and further linking (down to 160 contigs) has proceeded through joining of cosmid contigs via YAC bridges (1 and Coulson personal communication).

A major limitation in the application of YAC cloning to physical mapping has been the difficulty in obtaining purified YAC DNA in large amounts. At present the only means available

for purifying the YAC DNA from host chromosomal DNA requires pulsed field electrophoresis and yields are low. In the nematode project, for instance, YAC bridges had to be identified via hybridization between the pulse-field gel purified YACs and the already grouped cosmids. This procedure itself has been limited by the fact that YAC probes containing repetitive sequences in their genomic insert hybridize to several colonies on the library grids filters. Another limit is that YAC-YAC overlaps cannot be detected directly.

A method possibly circumventing both limitations is the use of repeated sequences to generate, via Southern hybridization, a fingerprint for each clone.

This approach allows the direct comparison of different YACs without the necessity of purifying the YAC DNA away from the yeast chromosomes, leading to the identification of overlapping clones. Restriction fragments containing members of each repetitive family are detected by hybridization and autoradiography, and the size, the level of hybridization and the number of different repetitive probes that give a positive signal are compared. YACs containing repetitive DNA sequences which were mapped this way form individual YAC contigs and can be included in the physical map.

MATERIALS AND METHODS

A genomic DNA library in yeast artificial chromosomes was prepared as previously described (1). The vector was pYAC4 and the *Saccharomyces cerevisiae* host strain was AB1380 (MAT a, y⁺, ura-3, trp-1, ade2-1, can-1 100, lys 2-1, his-5). The YAC clones to be used were selected from a library of approximately 2000 YACs with an average insert size of 250 kb, because they hybridized to multiple *C. elegans* cosmids.

Yeast DNA miniprep: 8 ml yeast cultures were grown in SD medium. DNA was prepared by the 'mini-prep' procedure described in Sherman et al. (3).

The DNA size markers on gels were a mixture of bacteriophage lambda DNA digested with HindIII, Eco RI and HindIII-EcoRI or alternatively a mixture of lambda DNA digested with HindIII, BglII and HindIII-EcoRI.

Southern transfer: was performed as described by Southern (4) with the following modifications: the DNA (1 µg of total yeast

* To whom correspondence should be addressed

DNA per lane) was transferred to Hybond-N membrane and immobilized by UV treatment. Hybridizations were performed in $6\times\text{SCP}$ ($20\times\text{SCP} = 2\text{M NaCl}$, $0.6\text{M Na}_2\text{HPO}_4$, $0.02\text{M EDTA pH } 6.2$), 0.9% Sarkosyl DS, 10% dextran sulphate and $100\mu\text{g/ml}$ single strand DNA carrier at 65°C for 6–16 hours. We used around 10^6 cpm of oligolabelled DNA probes (5,6) with a specific activity of greater than 10^9 cpm/ μg per hybridization bag. Labelled bacteriophage lambda DNA was included in the hybridization mix to detect the size markers. Following the hybridization, filters were washed four times in 1% SDS, $1\times\text{SCP}$ at 50°C and exposed on Fuji X-ray films with intensifying screens (Dupont Cronex Plus) for 16 hours at -70°C .

Removal of probe: In order to reuse the filters, after each independent hybridization, probes were removed from filters by repeating the following steps twice: 10 min r.t. in 0.2M NaOH , 0.5M NaCl , 10 min r.t. in $0.5\text{M Tris pH } 7.4$, 0.5M NaCl and 10 min r.t. in $1\times\text{SSPE}$ (3M NaCl , $0.2\text{M NaH}_2\text{PO}_4\cdot\text{H}_2\text{O}$, 0.02M EDTA , $\text{pH } 7.4$). Efficiency of removal was often monitored by re-exposing used filters before reuse.

Cosmids and Plasmids used as probes were transformed and maintained in *E. coli* rec A strain HB101 (F^- , hsd S20, (r^{BmB}), recA13, ara-14, proA2, lacY1, galK2, rpsL20 (Sm^r), xyl5, mtl-1, supE-44, l^-) selecting for ampicillin resistance. The strain was checked periodically for the recA phenotype.

Plasmid pRcA1, pRcD1, pRc35 (kindly provided by Gino Naclerio, this laboratory) and pRcC9 (7) were constructed in pGEM3/4, pRcB1 and pRcS5 were constructed in pEMBL8 (8). They each contain a member of a distinct DNA repetitive family.

RESULTS

The probes we used for fingerprinting YAC clones are from a collection of repetitive sequences which have previously been identified and shown to be widely distributed in the *C. elegans* genome (Naclerio et al. manuscript in preparation). These repetitive DNA families do not show high levels of restriction fragment length polymorphisms between different *C. elegans* strains. They have a copy number of between 30 and 150 per haploid genome and members have an average size between a few hundred and a few thousand base pairs, suggesting that in toto they cover 0.4% of the genome. We expect an element belonging to one of these six families on average every 250 kb.

These probes were tested against a subset of 159 YACs with an average insert size of 225kb.

Our sample of YACs was selected, from a genomic YAC library of 2000 clones (five genome equivalent), for the presence of repetitive sequences, by hybridization to cosmids known to contain uncharacterized repetitive DNA. We used about half of the positive YAC clones and thus estimate that this subset cover about 2.5 fold the selected fraction of repetitive DNA. Both previously unmapped and mapped YACs were included in the sample so that not only would we be able to detect new overlaps, but we would also be able to check the efficacy of the approach in detecting known overlaps (see table 1).

Each filter, containing the yeast DNAs, was hybridized in sequence with all the six repetitive probes. We found that each YAC was positive with at least one probe. Visual inspection of the films proved satisfactory for this small data set (159 YACs with six probes or 954 lanes) for recognizing bands in common between different clones. Examples of the Southern blots are shown in figure 1a and 1b, each of which includes two clones exhibiting

Fig. 1: Individual yeast colonies carrying YACs were grown, total DNA extracted and restricted with Eco RI. YACs sharing bands in common with RcD1 probe are indicated with arrows in fig. 1a and 1b. DNA size marker (M) was a mixture of bacteriophage lambda DNA digested with HindIII, BglII and HindIII-EcoRI. The overlap shown in fig 1a (three bands in common with this probe in the two lanes) was confirmed by the hybridization with a RcC9 probe on the same filter where other two bands common to the same YAC clones were identified (not shown).

Fig. 2a: *C. elegans* physical map in the area of chromosome V where the largest group of YACs (Y51C12 = Y56D6 = Y50A6 = Y54C8 = Y52A4 = Y44H1 = Y43E10) we joined together lies. Fig. 2b shows another group of YACs, on chromosome II already on map (Y54A10 = Y53F4 = Y56F7 = Y38B4) whose overlaps are identifiable by our method. The map figures were kindly provided by Alan Coulson. At the top left is shown the linkage group, shorter horizontal lines represent clone lengths of cosmids and their relative positions show the degree of overlap, as determined by fingerprinting, the longer lines represent YAC clones (names in brackets) with their overlap with cosmid clones (see reference 1)

Table 1.

	Characterized YACs	Uncharacterized YACs	TOT
number of clones:	94	65	159
overlapping clones detected:	17 (6 contigs)	11(3 contigs)	28 (9 contigs)
overlapping clones missed:	31 (13 contigs)	n.d.	n.d.
Size and description of the detected contigs:			
YAC denomination			LG
Y54A10 = Y53F4 = Y56F7 = Y38B4			II
Y52D1 = Y43E11 = Y39C11			n.d.
Y51D5 = Y37E10			V
Y55B12 = Y55C5 = Y41G1			n.d.
Y41D11 = Y41E3			IV
Y37E11 = Y40G3 = Y54E9			IV
Y43B2 = Y40G11			III
Y43A5 = Y44E11			II
Y51C12 = Y56D6 = Y50A6 = Y54C8 = Y52A4 = Y44H1 = Y43E10			V

The top part of the table compares the scores obtained for the two different sets of analyzed Yacs (previously mapped versus, not previously analyzed). The bottom part of the table shows the YAC contigs obtained with the 'repeat-fingerprint' method. Previously positioned on the physical map are indicated in bold face

bona fide overlaps. *In toto* we found 28 overlapping clones for a total of 9 contigs. A complete list of the observed overlaps is shown in table 1 (bottom part), where YACs previously

positioned on the physical map are indicated in bold face type. These previously positioned YACs allowed us to confirm several overlaps immediately. Figure 2a shows the physical map in the

Fig. 3: Y39C11, Y52D1 and Y43E11, indicated as a, b and c, digested with XbaI, with BamHI, with HindIII and again with EcoRI and hybridized with RcD1. Central lane (m) is the marker used that is a mixture of bacteriophage lambda DNA digested with HindIII, BglII and HindIII-EcoRI.

area of chromosome V where the largest group of YACs (Y51C12 = Y56D6 = Y50A6 = Y54C8 = Y52A4 = Y44H1 = Y43E10) we joined together lies. Figure 2b shows another group of YACs also already on the map (Y54A10 = Y53F4 = Y56F7 = Y38B4) whose overlaps are identifiable by our method.

In some cases of overlap, we did not have all the corresponding YACs positioned on the map. To verify the overlaps we proceeded with a further set of restriction digests with different enzymes. For example fig 3 shows three YACs run side by side in a control Southern experiment. Each YAC was digested with EcoRI, with HindIII, with XbaI and with BamHI: we observed bands in common in all the digests between the three YACs using two of our repetitive probes.

The clones from our sample which had been previously positioned on the physical map were also used to evaluate the efficiency of overlap detection of the method (see table 1, top part). Of the 94 clones which were already mapped, 48 clones (falling into 19 contigs) showed some overlap on the physical map. Of these 48 overlapping clones the repetitive fingerprinting method detected 17 (comprising 6 contigs). Thus about a third of all potential overlaps were detected by this procedure.

DISCUSSION

Repetitive sequences provide one means to generate a 'fingerprint' of YAC inserts. We monitored the successfulness of this mapping method, the potential applications and its limits using *C. elegans* genome as model system. The method is straightforward, utilizing standard technology. The analysis of each YAC takes little time and with some minor modifications the procedure could be

automated to allow processing of large numbers of clones. Of course with larger numbers of clones reading of the films and recognition of matches could be greatly facilitated through the use of digitized images and computer matching algorithms such as have been used in the cosmid fingerprint analysis of the *C. elegans* mapping project (9). Individual probes used singly provides valuable information over a pool of probes used together and in the present case allowed us to find overlaps with greater confidence.

A major drawback of the procedure is that a repetitive element homologous to one of the used probes is needed in the region of overlap between contiguous YACs or they will not appear related.

Although a useful number of overlaps were detected, the analysis of the clones already physically mapped by other means suggests that only 35% of all the potential overlaps were found (see table 1). This is not surprising since the repetitive families we used as probes in our fingerprinting experiments cover approximately only 5% of the repetitive fraction in *C. elegans*. On the other hand, a higher rate of overlap detection may be obtained when mapping genomes with a larger fraction of repetitive DNA (and a short period interspersed organization) or where repetitive families with a higher copy number proportionally to the genome size are present as suggested by Wada (10). These authors use repetitive fingerprinting to align some YACs known to contain DNA inserts from the human X-chromosome. We suggest, in general, that our approach to align YACs will be practical in higher eukaryotes, where shorter repeated elements such as (CG)_n or (CA)_n might be employed for indirect mapping in conjunction with traditional repeated

elements such as L1 and AluI (10). Unfortunately (CA)_n elements, which occur in mammals at a frequency of every 10–50kb, are present in all eukaryotic genomes including yeast and therefore they might be less useful than other simple sequences such as (CG)_n repeats, less frequent than the previous but not present in yeast (11). Even in mammalian genomes however a certain fraction of overlaps will go undetected because of an unfavorable distribution of repetitive sequences.

Furthermore, our method might become useful also for chromosome walking studies in organisms with a 'long period interspersion' genome, such as *Drosophila melanogaster* (12), where each time a repetitive region is found, the method could be used to bridge to the next unique area on the chromosome.

The genome of nematode *C. elegans* provided us with a rapid means of testing the approach. Because the physical map is nearing completion, we were able to evaluate the quality of the data. The genome is small enough that relatively few clones are required before significant numbers of overlaps occur. Similarly for others wishing to evaluate mapping approaches, the nematode genome may prove to be of use.

ACKNOWLEDGEMENTS

We thank John Sulston and Alan Coulson for suggestions and help and for sharing their data with us; John Pulitzer and Paul Harrison for critical reading of the manuscript, Gino Naclerio for plasmids pRcA1, pRcD1, pRc35 and Teresa Vespa and Rita Vito for technical help. This work was supported in part by CNR Progetto Finalizzato Biotecnologia, by CNR Progetto Speciale Bioinformatica and by USPHS grant GM41422.

REFERENCES

1. Coulson, A., Waterston, R., Kiff, J., Sulston, J., Kohara, Y. (1988) *Nature* **335**, 184–186.
2. Anand, R., Villasante, A., Tyler-Smith, C. (1989) *Nucl. Acids Res.* **17**, 3425–3433.
3. Sherman, F., Fink, G. R., Lawrence, C. W., (1974) *Methods in Yeast Genetics*. (Cold Spring Harbor, New York; Cold Spring Harbor Laboratory)
4. Southern, E. M. (1975). *J. Mol. Biol.* **98**: 503–517.
5. Feinberg, A. P., Volgenstein, B. (1983) *Analyt. Biochem.* **132**, 6–13.
6. Feinberg, A. P., Volgenstein, B. (1984) *Analyt. Biochem.* **137**, 266–267.
7. La Volpe, A., Ciaramella, M., Bazzicalupo, P. (1988) *Nucl. Acids Res.* **16**, 8213–8231.
8. Dente, L., Cesareni, G., Cortese, R. (1983) *Nucl. Acids Res.* **11**, 1645–1655.
9. Sulston J., Mallett, F., Staden, R., Durbin, R., Horsnell, T., Coulson, A. (1988), *CABIOS* **4**, 125–132.
10. 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**, 93–106.
11. Hamada, H., Petrino, M.G., Kakunaga, T. (1982) *Proc. Natl. Acad. Sci.* **79**, 6465–6469.
12. Davidson, E. H., Hough, B. R., Amenson, C. S. and Britten, R. J., (1973) *J. Mol. Biol.*, **77**:1.