Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors

David L.Neil[+], Alfredo Villasante[§], Richard B.Fisher, David Vetrie[φ], Brian Cox[1] and Chris Tyler-Smith*
Departments of Biochemistry and [1]Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Received December 22, 1989; Revised and Accepted February 22, 1990

ABSTRACT

The suitability of yeast artificial chromosome vectors (YACs) for cloning human Y chromosome tandemly repeated DNA sequences has been investigated. Clones containing *DYZ3* or *DYZ5* sequences were found in libraries at about the frequency anticipated on the basis of their abundance in the genome, but clones containing *DYZ1* sequences were under-represented and the three clones examined contained junctions between *DYZ1* and *DYZ2*. One *DYZ3* clone was quite stable and had a long-range structure corresponding to genomic DNA. All other clones had long-range structures which either did not correspond to genomic DNA, or were too unstable to allow a simple comparison. The effects of the transformation process and host genotype on YAC structural stability were investigated. Gross structural rearrangements were often associated with re-transformation of yeast by a YAC. *rad1*-deficient yeast strains showed levels of instability similar to wild-type for all YAC clones tested. In *rad52*-deficient strains, *DYZ5* containing YACs were as unstable as in the wild-type host, but *DYZ1/DYZ2* or *DYZ3* containing YACs were more stable. Thus the use of *rad52* hosts for future library construction is recommended, but some sequences will still be unstable.

INTRODUCTION

Tandemly repeated DNA sequences are widespread but poorly understood constituents of eukaryotic genomes (1). Their abundance, location in functionally significant regions of the chromosome such as centromeres and telomeres, and hypervariability make their analysis important for an understanding of genome organisation and evolution. Conventional cloning and sequencing experiments, complemented by genomic analysis using standard and pulsed-field gel electrophoresis, have provided a basis for the understanding of the structures of some of these sequences, while *in situ* hybridisation has provided information on their gross chromosomal localisations. However, some sequences have been difficult or impossible to clone satisfactorily in bacterial vectors (2). Furthermore, a full understanding of the long-range structure of the repeated region and its flanking sequences, and manipulation of these structures in model chromosomes in order to elucidate their biological activity, require cloning of longer fragments than can be accomodated in bacterial vectors. Yeast artificial chromosome vectors (YACs) offer a novel approach to these problems since they provide the possibility of cloning large stretches of DNA in a eukaryotic environment which is easily manipulated (3). We have therefore investigated the suitability of YACs for cloning tandemly repeated sequences.

The human Y chromosome is a good model system for the analysis of tandemly repeated sequences because it is rich in tandem repeats and because some of them have been relatively well characterised. Four abundant classes of repeat have been described, and their locations and structures are summarised in figure 1. *DYZ1* and *DYZ2* (4) are major constituents of the long arm heterochromatin. They have periodicities of 5bp and ~3.5kb (*DYZ1*) (ref 5) and ~2.5kb (*DYZ2*) (ref 6) and occupy many millions of base pairs (mb) of DNA. *DYZ3* (alphoid satellite DNA) is present at or close to the centromere. It has periodicities of ~170bp and ~5.7kb, and array sizes between ~250kb and ~1400kb have been observed (7 and unpublished observations). *DYZ5* (= *DYZ4*—see Materials and Methods) is located in the proximal region of the short arm. It has a periodicity of ~20kb and the size of the major array is between ~360kb and ~800kb long (8 and unpublished observations).

We have asked whether clones containing these sequences are found in YAC libraries, and whether their structures correspond to the genomic structures. We find that some such clones are present, but in all cases we find structural instability. Detectable rearrangements can be as frequent as one per two generations.

MATERIALS AND METHODS

DNA probes and hybridisation conditions

The *DYZ1* probe λHS5 (5) and the *DYZ2* probe pHY2.1 (6) were gifts from Howard Cooke, MRC Human Genetics Unit,

* To whom correspondence should be addressed

[+] Present addresses: [+]50 Chilswell Road, Oxford, UK, [§]Centro de Biologia Molecular (CSIC-UAM), Cantoblanco, 28049 Madrid, Spain and [φ]Pediatric Research Unit, Guy's Tower, London Bridge, London SE1 9RT, UK

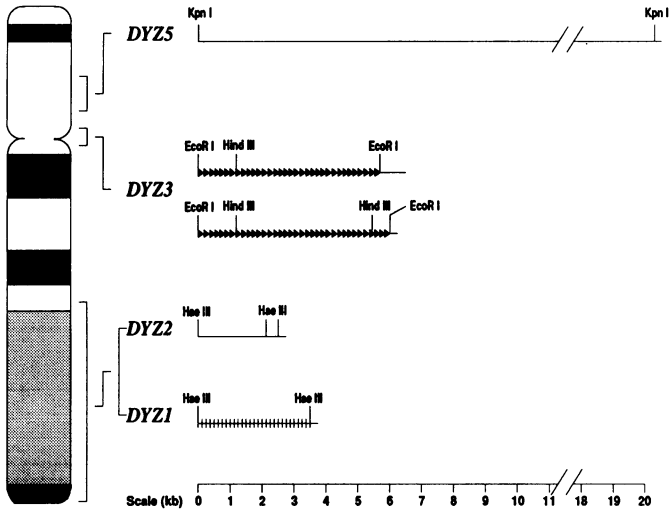

Figure 1. Summary of the location and structure of the mayor Y chromosome tandemly repeated sequences. *DYZ1* and *DYZ2* are located mainly in the long arm heterochromatin. *DYZ1* has a unit of ~3.5kb which contains 1 HaeIII site and is made up of 5bp subunits. *DYZ2* has a unit of ~2.5kb which contains 2 HaeIII sites. *DYZ3* is located close to the centromere and consists of ~5.7kb units containing 1 EcoRI site and 1 HindIII site, and a small number of ~6.0kb units containing 1 EcoRI site and 2 HindIII sites. Both units are made up of ~170bp subunits. *DYZ5* is located in the proximal region of the short arm. It consists of ~20kb units which contain 1 KpnI site.

Edinburgh. The *DYZ3* probe was pYαI, a 6.0kb unit of Y alphoid DNA subcloned from cY77 into pEMBL18$^+$ as an EcoRI–EcoRI fragment (7). The *DYZ4* probe pDP105 was purchased from the American Type Culture Collection NIH repository of human DNA probes and libraries. This sequence was found to produce a hybridisation pattern similar to that of the 4.5kb subclone of cY91 (a *DYZ5* sequence) when used to probe cY35 (8) and HindIII or XbaI digests of human DNA. It was therefore considered that *DYZ4* and *DYZ5* were the same locus. The *DYZ5* probe was a mixture of 91H6.4, 91H4.5, 91H3.5 and 91H2.2: respectively, the 6.4kb, 4.5kb, 3.5kb and 2.2kb HindIII fragments of cY91 subcloned in pTZ18R (8).

All probes were hybridised in 6×SET (1×SET = 150mM sodium chloride, 20mM Tris-HCl, 1mM EDTA, pH 7.8), 6×Denhardts (1×Denhardts = 0.02% (w/v) bovine serum albumen, 0.02% (w/v) polyvinylpyrrolidone and 0.02% (w/v) ficoll), 10% (w/v) dextran sulphate, 0.1% (w/v) sodium pyrophosphate, 0.1% (w/v) sodium dodecyl sulphate and 200μg/ml denatured sonicated salmon sperm DNA. *DYZ1*, *DYZ2*, *DYZ5* and pBR322 were hybridised at 68°C; *DYZ3* was hybridised at 79°C. Filters were washed in 0.1×SSC (1×SSC = 150mM sodium chloride, 15mM sodium citrate) at 65°C. Other procedures were as described previously (7, 8).

YAC library construction and screening

Two YAC libraries were used. Both contained inserts of human 49,XYYYY DNA (9) from the cell line OXEN. An SstI library was constructed in a modified version of the vector pYAC4 (ref 3) containing a polylinker in place of the EcoRI site; the polylinker included an SstI site. This vector was provided by William Brown and Andy Porter, Department of Biochemistry, Oxford. High molecular weight human DNA prepared in agarose plugs at 10^7 cells/ml was digested to completion with SstI (BRL) using 20 units of enzyme per ~2.5μg DNA for 4 hours. The plugs were melted, ligated to vector cut with SstI and dephosphorylated at an estimated molar ratio of insert: vector of 1: 10, and used to transform *Saccharomyces cerevisiae* AB1380 (*MATα* ura3 trp1 ade2.1 can1.100 lys2.1 his5 [ψ+]; a gift from M. V. Olson, Department of Genetics, Washington School of Medicine, St Louis.) according to the protocol of Burgers and Percival (10). Transformants were selected on plates lacking uracil and tryptophan. 30 plates containing about 3000–4000 colonies each were prepared. The colonies on a plate were mixed and about 10 000 cells from each pool were replated on a 20cm×20cm nitrocellulose filter. A replica of the filter was screened by standard colony hybridisation methods (11).

The second library was constructed from a partial EcoRI digest of OXEN DNA in the vector pYAC4 using PFGE fractionation as described (12). Colonies were picked individually from the transformation plates which lacked uracil onto plates lacking both uracil and tryptophan, and red colonies were stored in 96-well microtitre dishes. About 2 500 colonies derived from a series of trial experiments using different fractionation conditions were plated onto Immobilon-N PVDF filters (Millipore) using a 96-prong device as described by Brownstein *et al.* (13) and screened by colony hybridisation. The average insert size was about 200kb; thus the library was expected to contain about one third of a Y chromosome.

Colonies from either library which showed hybridisation were analysed without further colony purification in order to characterise the DNA insert after the minimum period of growth in yeast. Clones isolated from the SstI library are designated YOS and clones isolated from the EcoRI library are designated YOR.

Analysis of YAC clones

Yeast chromosomal DNA minipreps were prepared in agarose plugs by the lithium dodecyl sulphate method (12). Restriction enzyme digestions were performed as described previously (8). DNA samples were analysed by conventional gel electrophoresis or by pulsed field gel electrophoresis using the 'waltzer' apparatus (14). 1.5% agarose (Sigma type I) 'waltzer' gels were run in 0.5×TAE (1×TAE = 40 mM Tris acetate, 2mM EDTA) at 150V with a temperature of 20°C; pulse times and run times for individual gels are given in the figure legends.

Transfer of YACs to new hosts

In addition to *S. cerevisiae* AB1380, three repair-defective *S. cerevisiae* strains were used. They were 814-9/4c (*MATa ade2.1 trp1.1 his3.11 his3.15 leu2.3 leu2.112 ilv1.1 ura3.1 rad1::LEU2* [ψ+]), 814-9/2d (*MATa ade2.1 trp1.1 his3.11 his3.15 leu2.3 leu2.112 lys2.1 ura3.1 rad52::TRP1* [ψ+]) and 814-7/4c (*MATα* ade2.1 trp1.1 leu2.3 leu2.112 lys1.1 ura3.1 rad1::LEU2 rad52::TRP1 [ψ+]). These strains gave white colonies. They were derived from the following strains: W675-6a (*MATα* leu2.3 leu2.112 trp1.1 can1.100 ura3.1 ade2.1 rad52::TRP1 rad1::LEU2 [ψ+]; a gift from R. Rothstein), CM191/6a (*MATa ura3.1 can1.100 his5.2 lys1.1 trp5.48 SUP16* [ψ+]) and AB1380.

Chromosome I in AB1380 was ~230kb long, in the 814-9/4c (*rad1*) and 814-7/4c (*rad1, rad52*) strains it was ~240kb, and in 814-9/2d (*rad52*) it was ~250kb. 814-9/2d (*rad52*) and 814-7/4c (*rad1, rad52*) showed faint hybridisation to pBR322 DNA while AB1380 and 814-9/4c (*rad1*) did not. These characteristics ensured that no confusion of these four strains took place during analysis.

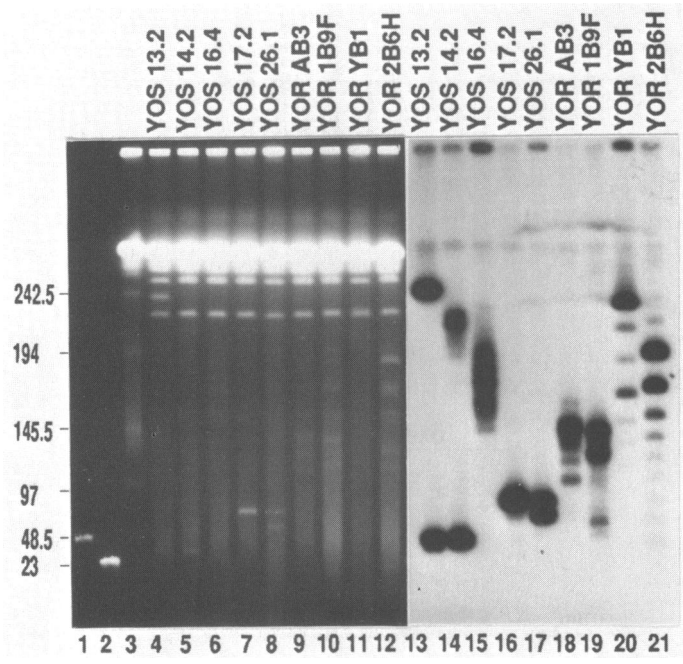

Figure 2. PFGE analysis of undigested YAC clones containing tandemly repeated DNA sequences. Tracks 1–12: photograph of an ethidium bromide stained gel (effective pulse time 10 sec, run time 32 hrs). Samples: 1, λ monomer; 2, λ HindIII; 3, λ multimers; 4–12, YAC clones as indicated at the top of each track. Tracks 13–21: filter transfer of lanes 4–12 probed with pBR322.

YACs were introduced into these strains by transformation (10) using total yeast DNA prepared in agarose plugs; the agarose was melted before transformation. The transformation efficiencies of the different strains measured using the modified pYAC4 or YAC-containing yeast DNA varied by about two-fold between strains. The genotypes of selected transformants were confirmed by measuring their radiation sensitivity.

RESULTS

Isolation of YAC clones containing human Y chromosome tandemly repeated DNA sequences

20 pools of YAC clones from the SstI library (about 70 000 colonies) were screened with the *DYZ1* probe. All pools contained clones which showed some hybridisation to the probe, but only three of the pools contained clones which originated from the Y chromosome as judged by their EcoRI and HaeIII digestion patterns. Cross-hybridisation to non-Y sequences is expected for this probe. Clones YOS 13.2, YOS 14.2 and YOS 16.4 are shown in figure 2 (tracks 4, 5 and 6).

30 pools of clones from the SstI library (about 100 000 clones) were screened with the *DYZ3* probe. Again, all pools contained weakly hybridising clones. However, only five pools contained clones which hybridised strongly and two of these are shown in tracks 7 and 8. The additional three clones contained a YAC similar to that in YOS 17.2, although two of the clones also contained a second, unrelated YAC (results not shown).

Screening the partial EcoRI library with *DYZ3* produced two clones (tracks 9 and 10). Screening with *DYZ5* also produced two clones (tracks 11 and 12).

In some of these clones an additional chromosomal band corresponding to the YAC is visible in the ethidium bromide stained gel photograph as expected (eg. track 7). In several clones, multiple bands are seen (eg. track 8); in others, no clear band is visible (eg. track 11). Hybridisation with pBR322 DNA reveals the YAC vector sequences and shows that vector is present in all clones (tracks 13–21). The clones which did not show a YAC band in the gel photograph contain a heterogeneous mixture of YACs. This heterogeneity could arise in several ways. Since the clones were not colony purified after isolation, they might be mixtures of different clones. Alternatively, the YAC in a single clone might be unstable and produce multiple bands during growth of the yeast culture. The experiments described in the next sections show that the lower bands in YOS 13.2 and YOS 14.2 are contaminants present in mixed colonies, but that all other examples of multiple bands are due to YAC instability.

Structure and instability of the *DYZ1/DYZ2* YACs

DYZ1 is characterised by a periodicity of ~3.5kb, and this is one of the most intense bands in OXEN DNA cut with HaeIII and probed with *DYZ1* (figure 3, track 2). Tracks 3–5 show that the three clones YOS 13.2, YOS 14.2 and YOS 16.4 all contain a fragment of this size. In addition, they all contain a fragment of ~6.5kb, and YOS 14.2 has a fragment of >23kb. Corresponding fragments are present in OXEN DNA. However, the *DYZ1* hybridisation was less intense than expected, suggesting that additional DNA sequences might be present in these clones. Since *DYZ2* is found near *DYZ1* in the genome (figure 1), the YAC clones were hybridised with *DYZ2*. Figure 3 (tracks 7–9) shows that all three hybridise. The patterns show some differences, but all of the YACs contain a strongly-hybridising ~2.1kb band and a weakly-hybridising ~6.5kb band. A ~6.5kb fragment also hybridised to *DYZ1* and thus probably contains a junction between the two sequences. All of the bands detected in the YAC DNAs are also observed in OXEN DNA.

Since these three clones are derived from a complete SstI digest, their long-range structure can be analysed by digestion with SstI and comparison with OXEN genomic DNA SstI fragments. Figure 3 (tracks 10–23) shows such an experiment. Undigested YAC DNA probed with *DYZ1* (tracks 11–13), or *DYZ2* (tracks 18–20) produces a hybridisation pattern similar to that produced by hybridisation to pBR322 (figure 2, tracks 13–15) except that the small contaminant YACs in YOS 13.2 and YOS 14.2 are not detected. This demonstrates that the heterogeneous high molecular weight species in YOS 14.2 and YOS 16.4 all contain vector, *DYZ1* and *DYZ2* sequences. Digestion with SstI decreases the size of the fragments detected by about 10kb, but leaves the patterns otherwise unchanged (tracks 14–16 and 21–23). This is the result expected from the method of construction of the library since about 10kb of vector sequences will be cleaved off each insert molecule. Examination of the OXEN hybridisation pattern produced by *DYZ1* in track 10 reveals a candidate sequence for the genomic copy of YOS 13.2 at about 230kb. However, reprobing of the filter (track 17) shows that all the fragments detected in OXEN DNA by *DYZ2* in this region of the gel are slightly different in size from the *DYZ1* fragment. Moreover, the candidate band in track 10 is similar in intensity to the YOS 13.2 insert in track 14. If the two bands contained the same sequence, the intensity of hybridisation in track 17 should be the same as the intensity in track 21. They are different. Therefore, despite its apparent stability, YOS 13.2 cannot be a true copy of a genomic fragment. The heterogeneity of YOS 14.2 and YOS 16.4 makes it

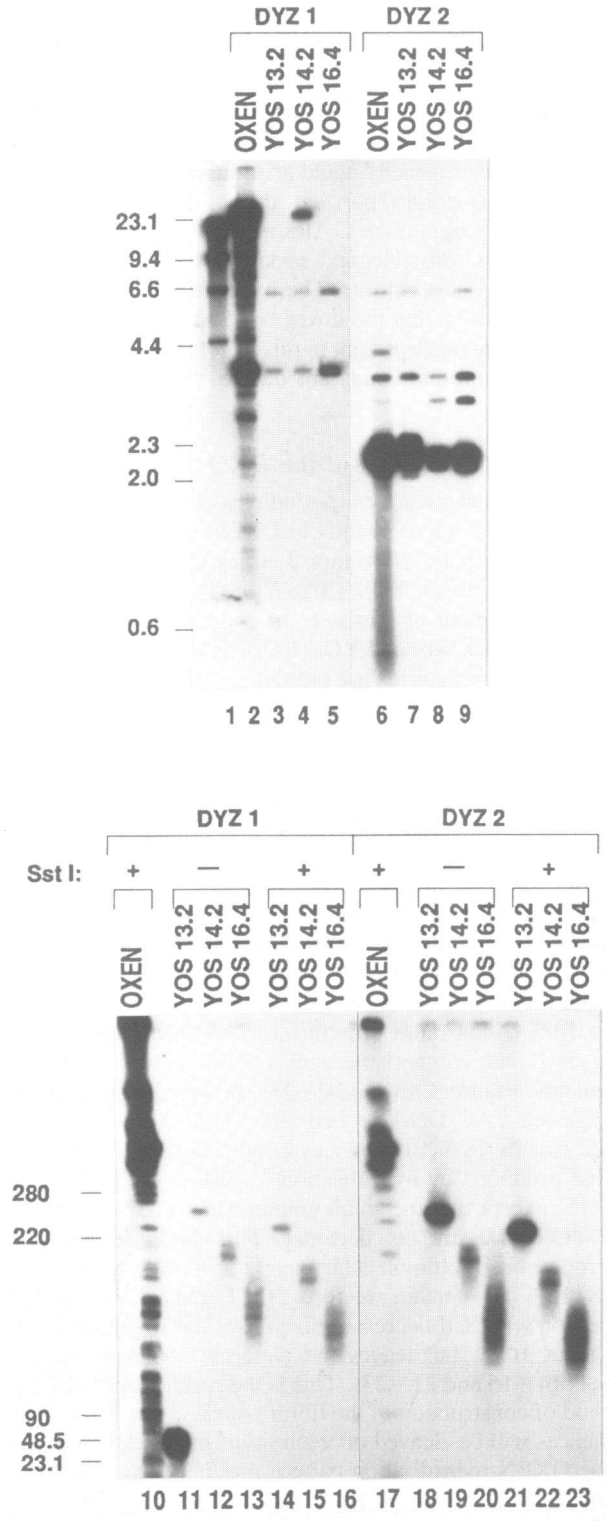

Figure 3. Structural analysis of the *DYZ1/DYZ2* YAC clones. Tracks 1–9: conventional gel electrophoresis. Samples: 1, λ HindIII; 2–9, HaeIII digests of the DNA indicated at the top of the track. Filter transfers were probed with *DYZ1* (1–5) or *DYZ2* (6–9). Tracks 10–23: PFGE analysis (effective pulse time 9 sec, run time 32 hrs). The identity of each DNA sample is indicated at the top of the track. Tracks 11–13 and 18–20 were undigested; tracks 10, 14–17 and 21–23 were digested with SstI. Filter transfers were probed with *DYZ1* (tracks 10–16), the radioactivity was allowed to decay for 4 months and the same filter was re-probed with *DYZ2* (tracks 17–23).

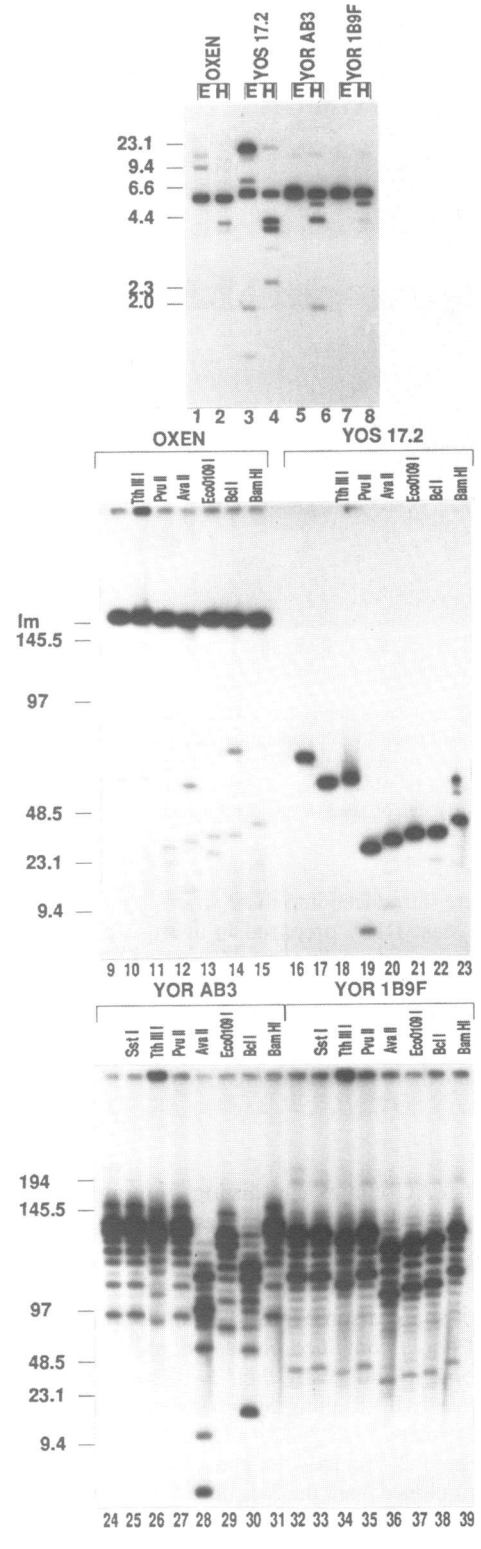

Figure 4. Structural analysis of the *DYZ3* (alphoid) YAC clones. All parts of the figure show filter transfers probed with *DYZ3*. Tracks 1–8: conventional gel electrophoresis of DNA samples digested with EcoRI (E) or HindIII (H) as indicated. Tracks 9–23: PFGE analysis of OXEN and YOS 17.2 DNA (effective pulse time 3 sec, run time 50 hrs). Track 16 is undigested; all other tracks are digested with SstI alone (9 and 17) or SstI plus the enzyme indicated at the top of the track. lm = limiting mobility position. Tracks 24–39: PFGE analysis of YOR AB3 and YOR 1B9F (effective pulse time 5 sec, run time 39 hrs). Samples are undigested (lanes 24 and 32) or are single digests with the enzyme indicated at the top of the track.

impossible to determine their origin by comparison with restriction fragments of genomic DNA.

Structure and instability of the *DYZ3* YACs

DYZ3 is organised in a predominant ~5.7kb unit which contains one EcoRI site and one HindIII site; a minority of units on the OXEN Y chromosome are ~6.0kb long and contain one EcoRI site and two HindIII sites generating ~4.1kb and ~1.9kb fragments upon digestion with HindIII (ref 7, and figure 4, tracks 1 and 2). Analysis of the YAC clones shows that all contain fragments in the ~5.7kb to ~6.0kb size range (figure 4 tracks 3–8). In EcoRI digests, most of the hybridisation to YOR 1B9F is at ~5.7kb, YOR AB3 has hybridisation at ~5.7kb and ~6.0kb, and YOS 17.2 has a more complex pattern with hybridisation at ~6.0kb and additional fragments which are both larger and smaller. In HindIII digests, YOR 1B9F shows hybridisation largely to ~5.7kb fragments, YOR AB3 has ~5.7kb, ~4.1kb and ~1.9kb fragments, and YOS 17.2 again has a more complex pattern including ~6.0kb fragments and several additional fragments. Thus YOR 1B9F is made up largely of ~5.7kb units, YOR AB3 is made up of ~5.7kb units and a few ~6.0kb units, and YOS 17.2 contains some ~6.0kb units and a large proportion of atypical units.

The long-range structure of YOS 17.2, like that of the *DYZ1/DYZ2* clones, can be analysed by digestion with SstI. Figure 4 (track 9) shows an SstI digest of OXEN DNA and track 17 shows an SstI digest of YOS 17.2. In the OXEN cell line the *DYZ3* array is about 1mb in size and contains a single internal SstI site ~65kb from one end (results not shown). Thus two fragments are seen in the SstI digest. YOS 17.2 produces a single fragment which corresponds in size to the ~65kb OXEN fragment. Tracks 10–15 show double digests of OXEN DNA in which the ~65kb SstI fragment is either not cleaved (track 10) or is cleaved (tracks 11–15). The ~60kb fragment in the AvaII digest (track 12), the ~20kb fragment in the EcoO109I digest (track 13) and the ~75kb fragment in the BclI digest (track 14) arise form cleavages elsewhere in the block. Tracks 17–23 show the corresponding double digests of YOS 17.2. In each case, the predominant fragment is the same size as a genomic fragment. A single strongly hybridising fragment (rather than two fragments) is seen in these YOS 17.2 digests because the *DYZ3* probe does not cross-hybridise under stringent conditions with the more diverged sequences from the edge of the array. Thus YOS 17.2 appears to be a relatively stable and unrearranged copy of a genomic DNA fragment. Three additional independent clones isolated from the same library had a similar structure (results not shown). The fifth independent isolate of this fragment, YOS 26.1, had a more complex structure and contained two bands (figure 2, tracks 8 and 17). Mapping experiments showed that the structure of the smaller YAC could be derived from that of the larger YAC by a single internal deletion event (results not shown).

The long-range structures of YOR AB3 and YOR 1B9F are analysed in figure 4 (tracks 24–39). Tracks 24 and 32 contain the undigested YACs; the other tracks are digests with restriction enzymes which do not cut frequently in the *DYZ3* array. In each track, a complex set of bands is visible, similar to those seen when pBR322 is used as the probe (figure 2, tracks 18 and 19). When both long and short exposures of the autoradiograph are examined, at least 19 bands can be counted in the YOR AB3 tracks, and 20 bands in the YOR 1B9F tracks, showing that these YACs are very unstable. The bands differ in size by about 6kb

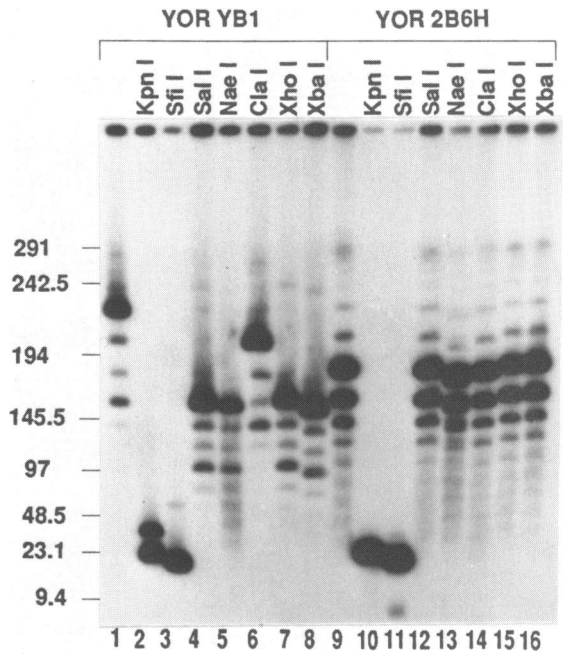

Figure 5. Structural analysis of the *DYZ5* YAC clones. DNA samples were analysed by PFGE (effective pulse time 10 sec, run time 32 hrs), filter transfer and hybridisation to a *DYZ5* probe. Tracks 1 and 9 are undigested. Tracks 2–8 and 10–16 are digested with the enzyme indicated at the top of the track.

(~ 1 unit) and vary in intensity. Digestion with restriction enzymes which do not cut within the YAC4 sequences (SstI, PvuII or BamHI) leaves the pattern unchanged. Digestion with enzymes which cleave the YAC4 sequences decreases the size of all bands by a constant amount. AvaII and BclI have additional cleavage sites within YOR AB3 (tracks 28 and 30). The presence of these sites allows the position of YOR AB3 within the OXEN array to be determined: it can be placed towards the opposite end from YOS 17.2 (unpublished results).

Structure and instability of the *DYZ5* YACs

DYZ5 is an array of ~20kb units, and the enzymes KpnI and SfiI cut once in most units (8). Digestion of YOR YB1 and YOR 2B6H with either of these enzymes produces a strong band of hybridisation at about 20kb (figure 5, tracks 2, 3, 10 and 11). The small difference in mobility between the KpnI band and the SfiI band may be due to the local difference in the amount of DNA in the track since KpnI and SfiI produce different fragment distributions, or to the presence of 2 or more SfiI sites close together in the units cloned. The KpnI digest of YOR YB1 has an additional band at about 40kb which may have arisen because some units in this YAC lack a KpnI site.

The other tracks in figure 5 show an analysis of the long-range structure of the *DYZ5* YACs. Tracks 1 and 9 contain undigested DNA. Both tracks have multiple bands of hybridisation resembling the pBR322 pattern (figure 2, tracks 20 and 21): at least 10 are present in track 1 and at least 15 in track 9. This demonstrates considerable instability. The bands vary in intensity and are spaced ~20kb apart (~ 1 unit). Five restriction enzymes which do not cut within the *DYZ5* array have been used to construct long range maps of genomic DNA (8). Digestion of YOR 2B6H with these five enzymes either has no detectable effect on the *DYZ5* hybridisation pattern (tracks 12, 15, 16) or reduces

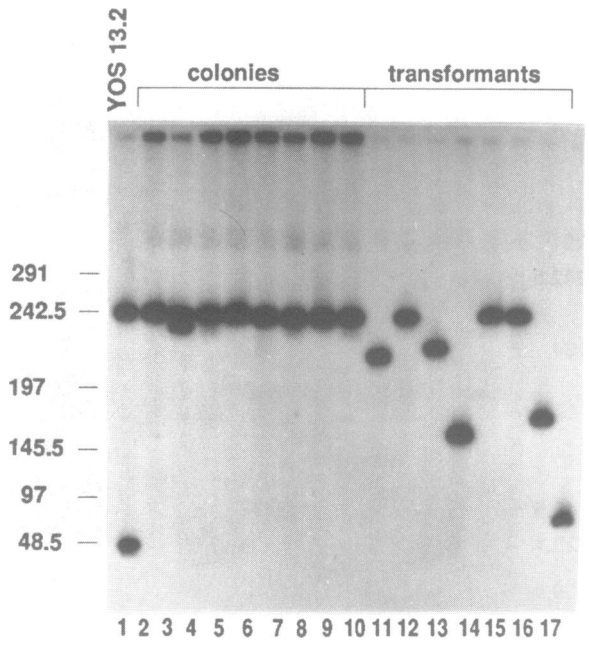

Figure 6. Effect of colony purification or re-transformation on YOS 13.2 DNA structure. Chromosomal DNA samples were analysed by PFGE (effective pulse time 11 sec, run time 32 hrs), filter transfer and hybridisation to pBR322. Track 1 is the original YOS 13.2 clone, tracks 2–9 are independent single colonies, and tracks 10–17 are independent transformants.

the size of all bands by a few kb (tracks 13, 14) due to cleavage within the YAC vector sequences. These results can be explained if the YOR 2B6H insert is derived entirely from within the *DYZ5* array. Digestion of YOR YB1 with these enzymes reduces the size of the bands by about 60kb (SalI, NaeI, XhoI and XbaI; tracks 4, 5, 7 and 8), or 20kb (ClaI; track 6). This suggests that YOR YB1 does not consist entirely of *DYZ5* sequence, but contains ~60kb of flanking sequence. These digestion patterns are consistent with the genomic structure and suggest that YOR YB1 spans the right-hand edge of the array: at the right-hand boundary there is a cluster of sites for SalI, NaeI, XhoI and XbaI, while the ClaI site occurs about 40kb into the flanking DNA (see figure 6 of ref 8).

Effect of transformation on YAC structural stability

Most of the YAC clones described above show evidence of structural instability. We wished to transfer these YACs into different yeast host strains in order to investigate the influence of host genotype on stability. The first experiment was therefore to determine the effect of the DNA transformation procedure itself on YAC structure. This can be done most easily using a YAC clone which is relatively homogeneous and stable in its original host. Figure 6 shows the results of an analysis of YOS 13.2. The culture was either streaked out and single colonies picked for analysis (tracks 2–9), or the DNA was used to re-transform AB1380, and single clones were analysed (tracks 10–17). The small contaminant YAC was not analysed in this experiment. Most colonies examined after re-streaking show no change in YAC structure. Only one colony (track 3) out of a total of 16 analysed shows a detectable change. In contrast, many of the clones produced by re-transformation show large changes in structure: in all, 9 out of 16 showed a detectable change. These YACs were not heterogeneous in size in a transformant. Thus

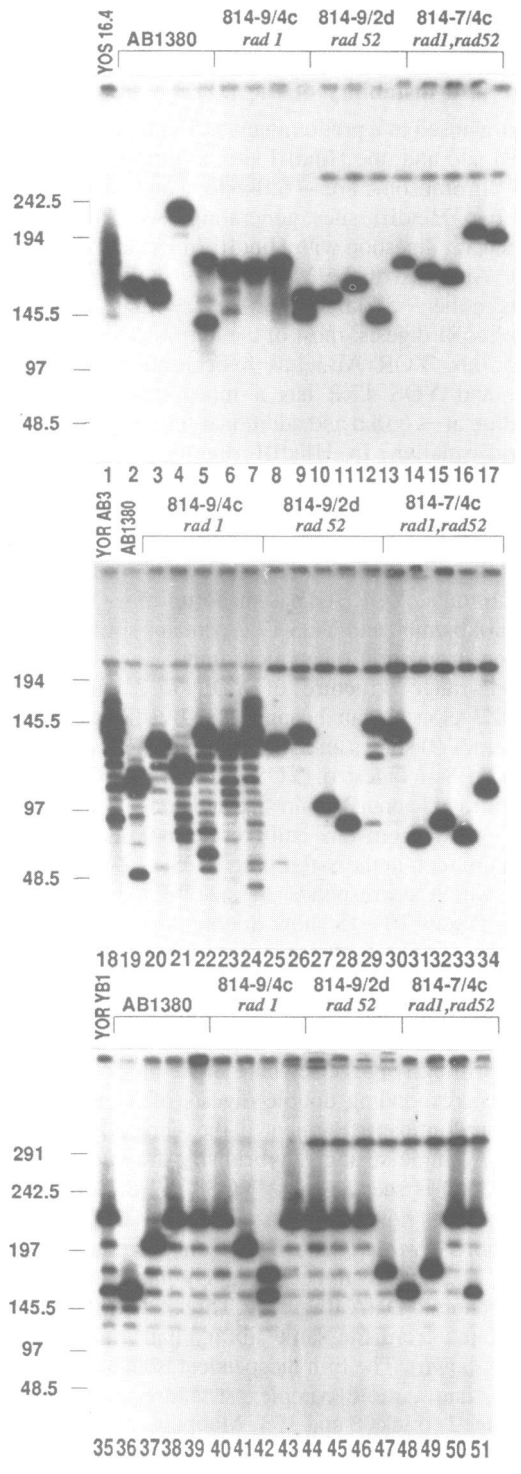

Figure 7. Effect of yeast genotype on YAC structural stability. All parts of the figure show filter transfers of chromosomal DNA samples run on pulsed field gels and probed with pBR322. Tracks 1–17: YOS 16.4 DNA (track 1) was used to transform AB1380 (tracks 2–5), 814-9/4c (*rad1*) (tracks 6–9), 814-9/2d (*rad52*) (tracks 10–13) or 814-7/4c (*rad1*, *rad52*) (tracks 14–17). Effective pulse time 8 sec, run time 35 hrs. Tracks 18–34: YOR AB3 DNA (track 18) was used to transform AB1380 (track 19), 814-9/4c (*rad1*) (tracks 20–24), 814-9/2d (*rad52*) (tracks 25–29) or 814-7/4c (*rad1*, *rad52*) (tracks 30–34). Effective pulse time 5 sec, run time 40 hrs. Tracks 35–51: YOR YB1 DNA (track 35) was used to transform AB1380 (tracks 36–39), 814-9/4c (*rad1*) (tracks 40–43), 814-9/2d (*rad52*) (tracks 44–47) or 814-7/4c (*rad1*, *rad52*) (tracks 48–51). Effective pulse time 11 sec, run time 32 hrs.

it appears that a single rearrangement event occured during about half of the transformations and that YACs were subsequently stable (like the original clone) during growth in yeast. In a similar experiment using YOS 17.2, one rearrangement was detected in 20 transformations, but no rearrangements were seen when either of two colonies was examined after growth for about 100 generations (results not shown). We assume that similar rearrangements can occur during the transformations described below, but are not readily detected because of the heterogeneity of the starting DNA.

Effect of yeast genotype on YAC structural stability

The instability detected in several YACs is likely to be mediated by the yeast host recombination apparatus. Two of the yeast genes which have major influences on ectopic (intrachromosomal) homologous recombination are *RAD1* and *RAD52* (15). We have therefore examined the effect of null mutations at these loci on the stability of unstable YACs. YOS 16.4, YOR AB3 and YOR YB1 DNAs were used to re-transform AB1380 or to transform the 814-9/4c (*rad1*), 814-9/2d (*rad52*) and 814-7/4c (*rad1, rad52*) strains. The results are shown in figure 7.

Transformants often contain YACs which differ in size from the starting material. This is expected both because single molecules are selected from a complex mix and because changes can occur during transformation (see above). YOS 16.4 re-transformants in AB1380 show heterogeneity within a colony (tracks 2–5). This is visible both as multiple distinct bands within a track, and as a smear of unresolved bands in some tracks. However, the heterogeneity is less than that in the original YOS 16.4 clone (track 1). Transformants in the 814-9/4c (*rad1*) host (tracks 6–9) are similar to those in AB1380. In contrast, all the transformants in a 814-9/2d (*rad52*) or 814-7/4c (*rad1, rad52*) background show only a single YAC band (tracks 10–17). Thus, the stability of YOS 16.4 (containing *DYZ1/DYZ2* sequences) is increased by the *rad52* mutation.

YOR AB3 transformants in AB1380 or 814-9/4c (*rad1*) hosts show considerable heterogeneity (tracks 19–24), similar to that in the original clone (track 18). Transformants in 814-9/2d (*rad52*) or 814-7/4c (*rad1, rad52*) hosts show reduced heterogeneity within a track (tracks 25–34). However, multiple bands are present in several tracks; they are most clearly seen in tracks 25, 29 and 30. Thus the *rad52* mutation also increases the stability of this *DYZ3* clone, but does not entirely eliminate instability.

YOR YB1 transformants in all four hosts (tracks 36–51) show levels of instability similar to the original clone (track 35). Thus for this *DYZ5* sequence, neither the *rad1* nor the *rad52* mutation significantly increases stability.

DISCUSSION

Representation of tandemly repeated DNA sequences in the YAC libraries

Are tandemly repeated sequences cloned in YAC vectors at the expected frequency? The expected representation of the *DYZ3* and *DYZ5* sequences in the partial EcoRI library can be estimated simply. For *DYZ3*, the block size is ~1mb in OXEN (unpublished observations). If the haploid genome is assumed to be 3000mb, and all sequences are cloned without bias, then since there are 2 Y chromosomes per haploid genome, 1 clone in 1500 should contain *DYZ3* sequences. For *DYZ5*, the block size is ~0.8mb (8) and the expected frequency of clones is 1 per 1900. The observed two clones from each locus out of ~2500 screened is in good agreement with the expected number.

The expected representation of sequences in the complete SstI library is more difficult to estimate. The human genome is expected to contain about 670 000 SstI fragments (16). If each YAC clone contained 1 SstI fragment and all fragments were equally represented, the ~100 000 clones in the library would contain about 1/6.7 genomes or about 1/3 Y chromosome. There is only one copy of the ~65kb *DYZ3* SstI fragment per chromosome. Thus there would be only about a 1 in 3 chance of finding this fragment. However, some clones contain more than one SstI fragment, and there is likely to be selection against the smallest fragments since small yeast chromosomes are mitotically unstable. These factors could account for the five independent copies of the *DYZ3* fragment obtained. Additional considerations apply to the representation of *DYZ1/DYZ2* sequences. In SstI digests of OXEN DNA, there are about 20 fragments between 50kb and 250kb in size which hybridise to *DYZ1* (figure 3, track 10). At most, 3 of these could be the same size as fragments which hybridise to *DYZ2* (figure 3, track 17). YAC clones were selected on the basis of their hybridisation to *DYZ1* alone. Therefore most of the clones picked would not be expected to hybridise to *DYZ2*. The observation that all of the three clones examined contain both *DYZ1* and *DYZ2*, as well as the relatively small number of *DYZ1* clones (compared to *DYZ3* clones and the number of genomic bands hybridising to each sequence) suggests that there is a bias against cloning *DYZ1* alone in YACs.

Factors affecting the stability of tandemly repeated sequences in YACs

Although structural instability could be detected in all of the YAC clones examined, the instability varied considerably between clones. YOS 13.2 and YOS 17.2 were sufficiently stable to allow a comparison of their structure with that of OXEN DNA, and YOS 17.2 was representative of a genomic DNA fragment (figure 4) while YOS 13.2 was not (figure 3). In contrast, the other clones were too unstable to allow such a simple comparison.

Which characteristics of the insert structure could account for these differences in stability? YOS 17.2 is the smallest of the clones, but YOS 13.2 is the largest. Thus size alone is not a major determinant of stability. One out of three *DYZ1/DYZ2* clones was relatively stable and one out of three *DYZ3* clones was also relatively stable. Thus the class of repeat is not an over-riding factor. YOS 17.2 contained particularly diverged Y alphoid repeat units (figure 4, tracks 3 and 4) and, if instability is due to recombination, and recombination depends on homology, this may make a major contribution to its stability. However, the repeat units of YOS 13.2 do not appear to be more diverged than those of YOS 14.2 or YOS 16.4 in their HaeIII site distributions (figure 3). The stability of YOS 13.2 cannot be satisfactorily explained by any single factor; a combination of factors may be important.

Host genotype can also significantly affect stability. A *rad1* host did not increase the stability of any clone. However, *rad52* hosts increased the stability of the *DYZ1/DYZ2* YAC and the *DYZ3* YAC, although they made little difference to the *DYZ5* YAC (figure 7). The finding of a difference between *rad1* and *rad52* hosts is not unexpected since it has been shown that different types of recombination events are controlled by the *RAD1* and *RAD52* genes (15). Previous work on *RAD1* has shown that although a *rad1* mutation reduced the frequency of

intrachromosomal recombination between duplicated alleles carrying deletions, it did not reduce the frequency of recombination between duplicated alleles carrying point mutations (17). Units in the human tandemly repeated arrays are expected to differ largely by point mutations.

The mechanism(s) responsible for instability are not known. There is a striking difference between the unit structure, which is relatively homogeneous within the YAC and always consistent with the genomic structure, and the long-range structure, which can be very heterogeneous and difficult to compare with the genomic structure. This suggests that most rearrangements are mediated by homologous recombination events which involve integral numbers of units. Simple intramolecular deletion events cannot fully account for the observed patterns because larger multimers are sometimes produced (figure 7). Unequal sister chromatid exchange could produce such larger molecules, but should produce equal numbers of larger and smaller molecules. It may be that a combination of the two mechanisms is at work, or that unequal exchange is generally accompanied by selection for the smaller molecule, or that other mechanisms are responsible for the observed patterns.

Rearrangement events are very frequent. In an experiment such as that shown in figure 7, about 30–35 generations elapse between transformation and analysis. Some of the samples shown in figure 7 contain over 20 different bands. If each of these is generated by a single rearrangement, the frequency of such events must on average be greater than one per two generations.

Implications for genomic analysis

The observation that the yeast transformation procedure itself can cause changes in YAC structure is striking. Such rearrangements could account for the fact that the insert of YOS 13.2 does not correspond to a genomic fragment. If rearrangements occur frequently in clones which do not contain tandemly repeated sequences, they could limit the usefulness of YAC cloning as a general method for analysing genomic structure.

Several reports have emphasised YAC stability and the similarity between sequences cloned in YACs and their genomic counterparts (3, 13), although long-range analysis has rarely been carried out. Nevertheless, other unstable YACs have been reported. Analysis of 454 YACs containing *Drosophila* DNA revealed six unstable clones (18); two of these contained ribosomal DNA, a tandemly repeated sequence. Analysis of 27 YAC clones containing human ribosomal DNA did not reveal instability (19), but none of the clones examined contained multiple copies of the tandemly repeated ribosomal DNA unit. It is not known whether such clones would be stable. Yeast chromosomes containing human telomeric tandem repeats cloned as functional telomeres have been found to have undergone large-scale deletion, although the final structures examined showed only limited heterogeneity (20). It thus seems likely that many tandemly repeated sequences will be unstable in YACs, but that some may no longer show heterogeneity by the time that sufficient DNA is available for analysis. On the basis of the results presented in this paper, it would seem worthwhile to use *rad52* hosts for future library construction. However, some sequences (such as *DYZ5*) will be unstable even in *rad52* hosts. Comparison of clone structure with genomic structure is always necessary, but is particularly important for these repeated sequences.

ACKNOWLEDGEMENTS

We thank Philip MacKinnon for help with library screening, and Ed Southern and William Brown for comments on the manuscript. This work was supported by the CRC. D.V. was supported by a Commonwealth Scholarship.

REFERENCES

1. Beridze, T. (1986) Satellite DNA. Springer-Verlag, Berlin.
2. Brutlag, D., Fry, K., Nelson, T. and Hung, P. (1977) *Cell* **10,** 509–519.
3. Burke, D. T., Carle, G. F. and Olson, M. V. (1987) *Science* **236,** 806–812.
4. Cooke, H. (1976) *Nature* **262,** 182–186.
5. Cooke, H. J. and Hindley, J. (1979) *Nucleic Acids Res.* **6,** 3177–3197.
6. Cooke, H. J., Schmidtke, J. and Gosden, J. R. (1982) *Chromosoma* **87,** 491–502.
7. Tyler-Smith, C. and Brown, W. R. A. (1987) *J. Mol. Biol.* **195,** 457–470.
8. Tyler-Smith, C., Taylor, L. and Müller, U. (1988) *J. Mol. Biol.* **203,** 837–848.
9. Sirota, L., Zlotogora, Y., Shabtai, F., Halbrecht, I. and Elian, E. (1981) *Clin. Genet.* **19,** 87–93.
10. Burgers, P. M. J. and Percival K. J. (1987) *Analyt. Biochem.* **163,** 391–397.
11. Sherman, F., Fink, G. R. and Hicks, J. B. (1986) Laboratory Course Manual for Methods in Yeast Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor.
12. Anand, R., Villasante, A. and Tyler-Smith, C. (1989) *Nucleic Acids Res.* **17,** 3425–3433.
13. Brownstein, B. H., Silverman, G. A., Little, R. D., Burke, D. T., Korsmeyer, S. J., Schlessinger, D. and Olson, M. V. (1989) *Science* **244,** 1348–1351.
14. Southern, E. M., Anand, R., Brown, W. R. A. and Fletcher, D. S. (1987) *Nucleic Acids Res.* **15,** 5925–5943.
15. Klein, H. (1988) *Genetics* **120,** 367–377.
16. Drmanac, R., Petrovic, N., Glisin, V. and Crkvenjakov, R. (1986) *Nucleic Acids Res.* **14,** 4691–4692.
17. Schiestl, R. H. and Prakash, S. (1988) *Mol. Cell. Biol.* **8,** 3619–3626.
18. Garza, D., Ajioka, J. W., Burke, D. T. and Hartl, D. L. (1989) *Science* **246,** 641–646.
19. Labella, T. and Schlessinger, D. (1989) *Genomics* **5,** 752–760.
20. Brown, W. R. A. (1989) *Nature* **338,** 774–776.