Complete coverage of the *Schizosaccharomyces pombe* genome in yeast artificial chromosomes

Elmar Maier, Jörg D. Hoheisel[1], Linda McCarthy[1], Richard Mott[1], Andrei V. Grigoriev[1], Anthony P. Monaco[2], Zora Larin[1] & Hans Lehrach[1]

The genome of the fission yeast, *Schizosaccharomyces pombe*, consists of some 14 million base pairs of DNA contained in three chromosomes. On account of its excellent genetics we used it as a test system for a strategy designed to map mammalian chromosomes and genomes. Data obtained from hybridization fingerprinting established an ordered library of 1,248 yeast artificial chromosome clones with an average size of 535 kilobases. The clones fall into three contigs completely representing the three chromosomes of the organism. This work provides a high resolution physical and clone map of the genome, which has been related to available genetic and physical map information.

[1]*Genome Analysis Laboratory, Imperial Cancer Research Fund P.O. Box 123, Lincolns Inn Fields, London WC2A 3PX, UK*
[2]*Human Genetics Laboratory, Imperial Cancer Research Fund Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK*

Correspondence should be addressed to E.M.

The DNA of *Schizosaccharomyces pombe* shows all the basic characteristics of a eukaryotic genome. This allows the comparative examination of important biological functions in a relatively simple model organism[1–4], as demonstrated by intensive studies on the cell cycle division process[5–8]. Due to this inherent biological interest and the availability of genetic[9,10] and physical mapping information[11], *S. pombe* serves as an excellent system to develop and test techniques for the mapping of entire chromosomes. An ordered yeast artificial chromosome (YAC) library provides a way to bridge the gap left between the level of linkage and cytogenetic mapping and conventional recombinant DNA technology. Also, the potential difficulties associated with the use of the YAC cloning system[12] could be addressed in a way that is difficult to achieve in the analysis of larger genomes.

Hybridization fingerprinting on high density filters should provide an efficient route to map large numbers of clones preferentially using probes detecting multiple sites in the genome[13]. However, since the genome of *S. pombe* is relatively small and the average size of the YAC clones is comparatively large (Fig. 1), the use of unique probes was very efficient, permitting a direct data analysis. The results presented here, together with the mapping on cosmid level we have embarked upon using various types of probes, will eventually allow an evaluation of the overall efficiency of the different methods.

Map generation

The *S. pombe* YAC library covering the genome about 47 times (see Methodology) was hybridized with 34 genetic markers, distributed over the whole genome, and also with a set of anonymous clones from our cosmid library (Fig. 2). Hybridization of entire cosmids onto YACs and *vice versa* does not require any competition, since the vector systems have no significant homology. Initially the cosmid probes were randomly selected. At a later stage of the analysis, a set of 65 YAC clones already ordered into contigs was hybridised to the cosmid library with the intention of both increasing the efficiency of probe selection and simultaneously subdividing the cosmid library. Cosmids not identified by any of these YAC clones, and therefore likely to be located in one of the

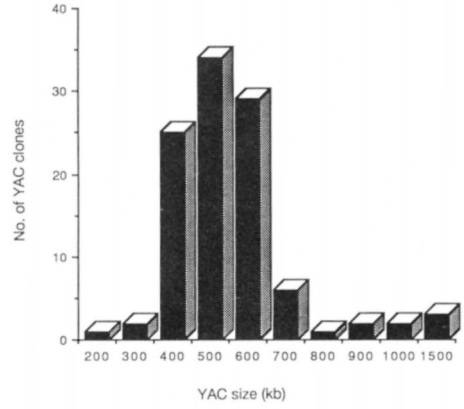

Fig. 1 Size distribution of the *S. pombe* YAC library. 71 randomly selected clones and 35 specific clones (total of 106 clones) were taken into account (8.5% of the library). The average clone size was 535 kb.

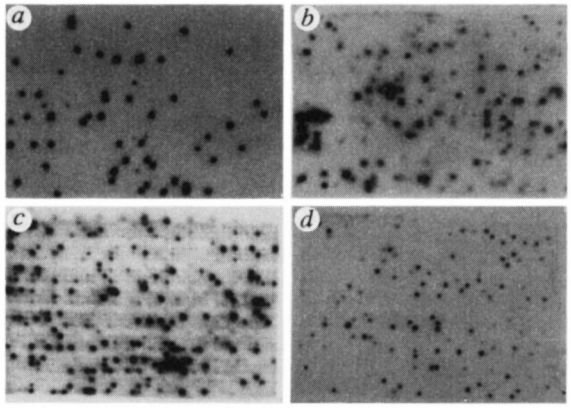

Fig. 2 Hybridization patterns of entire cosmids to high density gridded YAC filters. The cosmids contain: *a*, unique insert; *b*, centromeric elements[15]; *c*, ribosomal DNA[16]; *d*, retrotransposon[17]. Low background levels facilitate the assignment of the grid positions of the positive clones.

remaining gaps of the YAC clone coverage, were selected as probes to screen the YAC library. This proved to be an important prerequisite for an efficient ordering process.

In the analysis, more accurate and longer contigs were obtained by ordering the probes first and then fitting the clones to the probe order. Using newly designed algorithms based on simulated annealing and on heuristic rules, the probes were ordered by calculating the distances between probes on the basis of their hybridization frequencies to the YAC clones. This produced a set of probe tagged sites (PTS), named in analogy with the concept of a sequence tagged site (STS)[14]. Based on this high-density PTS map and using the same hybridization data, the YAC clones could be localized unequivocally, even those containing large repetitive elements like the centromeres[15], the main ribosomal DNA[16] and retrotransposable regions[17]. The large and complex centromeres spanning 40, 70 and more than 100 kilobases (kb) of repetitive sequence in the three chromosomes, respectively[15], and the location of about 100 copies of the 10.4 kb ribosomal DNA repeat[15] on both ends of chromosome III[18], are visible in Fig. 3 as high density patterns of positive signals outside the contig diagonal. However, a few copies of the rDNA repeat have also been detected on chromosome I and II.

Every part of the genome seems to be equally represented in our YAC library based on homologous distribution of the hybridization frequencies. There is no evidence that any region is uncloneable in YACs. With an initial probe density of about one in 200 kb, eight contigs were obtained. All of the five remaining gaps have been closed by isolating the insert ends of the YAC clones[23] positioned at the ends of contigs and using them as probes.

1,151 YAC clones out of 1,248 (92%) have been positively hybridized with a total of 153 probes. 4% of the library did not grow on the filters and the remaining clones contain small inserts falling between pairs of probes (not shown). In seven cases, the order of a pair of probes, which according to their similar hybridization patterns are located very close to each other, might be the reverse orientation. This does not affect the clone order, however.

Relation to genetic and physical mapping data

The map we have obtained is consistent with the genetic map[10] and *Not*I restriction map[11] (Fig. 4), although some minor contradictions to the recently published *Sfi*I restriction map[19] have been detected (not shown). These might be due to different strain sources.

We also detected some inconsistencies with previous genetic data. The linkage between *ade5* and *wee1* (ref. 20) could not be confirmed. According to the hybridization patterns *wee1* is located next to *cut1* (ref. 21)on the short arm of chromosome III. Also, the block of genetic markers stretching from *nda2* to *nda3* on the long arm of chromosome II is reversed so that *nda3* is closer to the centromere. Furthermore, one out of ten recently published DNA repair (rad) genes[22] has been located differently: *rad11* was found to be on the short arm of chromosome II in our map, a result confirmed by its hybridization to the *Not*I-fragment P(Fig. 4).

Frequency of chimaeric YAC clones

Chimaeric YAC clones, containing inserts derived from two nonadjacent fragments of DNA, are a problem in genome mapping. For the *S. pombe* library, the occurrence of such hybrids was analysed after ordering the clones by counting the frequency of positive hybridization signals outside the confirmed contigs (Fig. 3). 39% of the YAC clones showed such non-contiguous hybridization patterns. About two thirds, however, were due to strong background signals or no growth of the relevant colony on some filters. This leaves a figure of about 13% hybrid clones, less than recently reported for a human Xq24–q28 YAC library[24]. Further examination showed that about 75% of identical chimeras are the result of duplicate picking.

Discussion

Since we have both a high genome coverage in our YAC library and a high probe density, problems caused by dispersed repeated sequences in the genome were very rare. In fact, the use of a reasonable number of genetic markers as well as the linkage of other probes to the *Not*I restriction map[11] eliminated this problem. Nevertheless, validation by verification of restriction fragment sizes is most convincing. Therefore, 26 YAC clones were digested with *Not*I confirming the clone order of the YAC library and aligning the YAC clones exactly to the *Not*I physical map (Fig. 4) whenever possible. The YAC clones as well as filters containing either the complete library or the minimal set of clones covering the entire genome are freely available upon request.

Preliminary calculations based on the presented hybridization results indicate that, given a library of similar quality, roughly a ten genome coverage is necessary for an unambiguous reconstruction of the map.

The YAC map covering the entire genome of *S. pombe* is a step towards a fully integrated high-resolution physical and genetic map of the unicellular fungus based on smaller size cosmids and P1 clones[25], providing an even more immediate access to any segment of the genome. Such an integrated map is close to completion at the ICRF (J.D.H *et al.*, in preparation). Similar work on the completion of a cosmid map of *S. pombe* is in progress (D. Beach and M Yanagida; D. Beach, personnal commuinication).

These maps will not only provide a high resolution

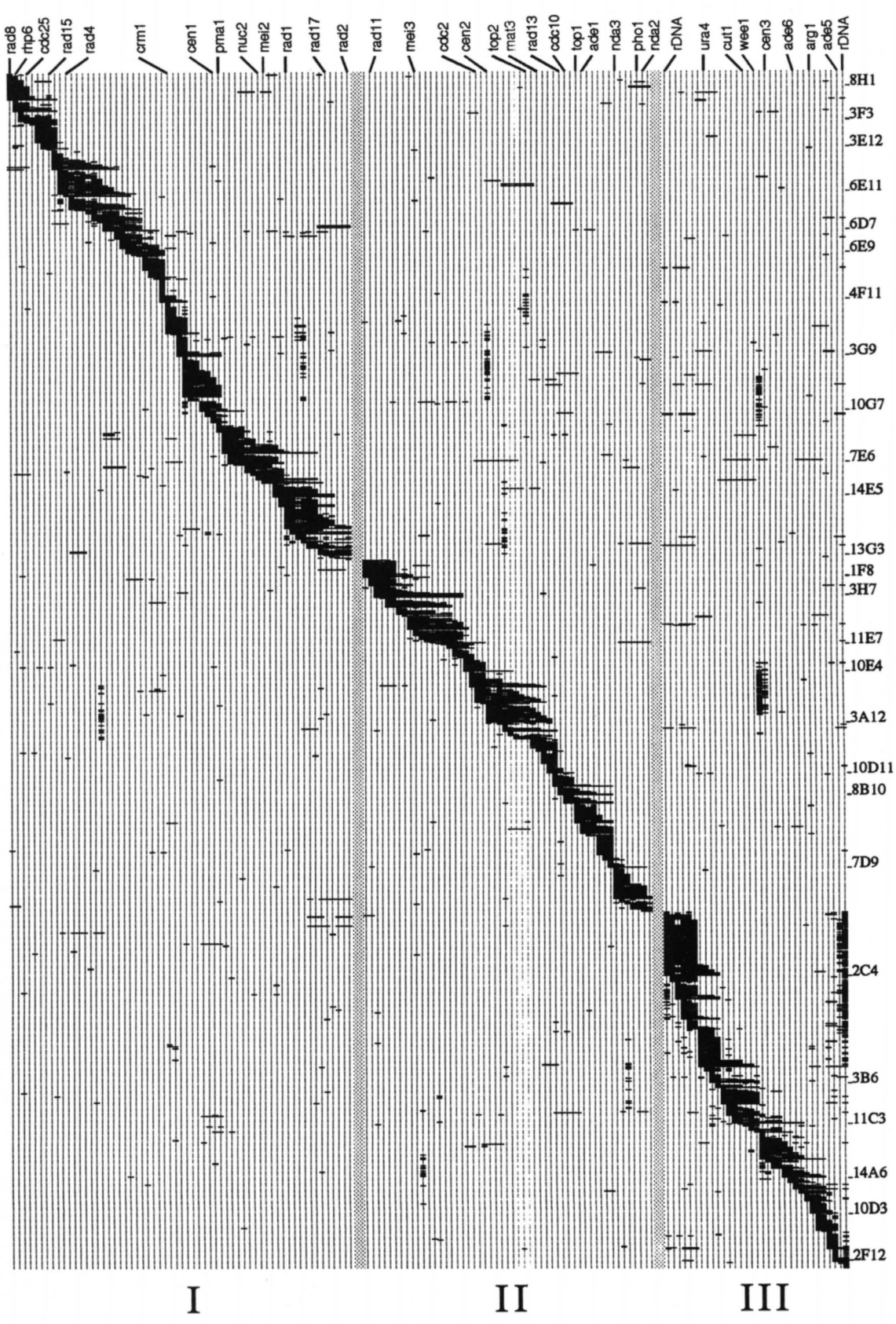

Fig. 3 The order of the clones from the *S. pombe* YAC library spanning the three chromosomes of the organism. YAC clones are shown on the vertical axis, where only the subset of 26 clones spanning the entire genome is indicated. The probes are arrayed on the horizontal axis, again only the genetic markers used are identified. The vertical grey bars separate the three chromosomes (I, II, III). The blocks of positive signal outside the constructed contigs indicate the locations of repeats, for example, the centromeres of the chromosomes, which are repetitive over 40, 70 and 100 kb, respectively[15]. Crosshybridization on account of the major ribosomal repeat, a 10.4 kb fragment[16] which occurs in total about 100 times on both ends of chromosome III[18] is also clearly visible. Retrovirus sequences with their long terminal repeat[17] are located on the short arm on chromosome I and on the long arm of chromosome II and III.

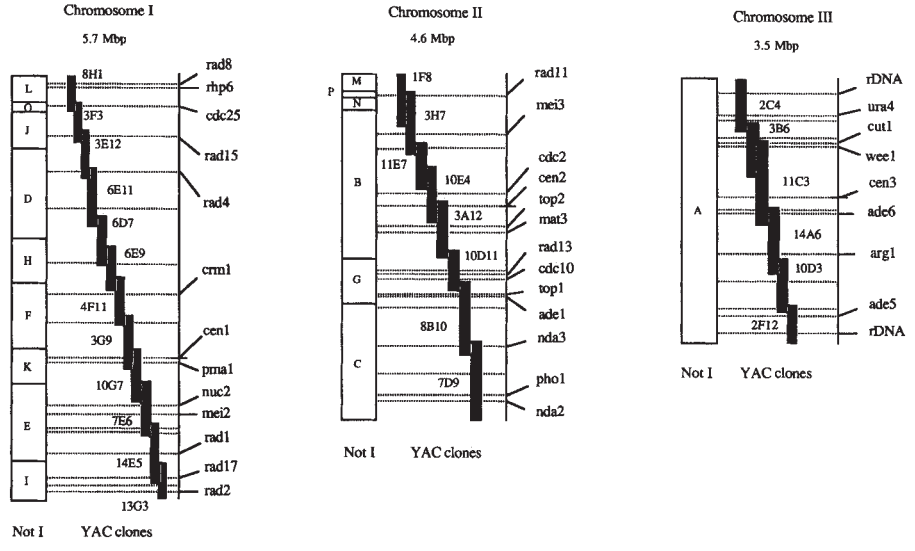

Fig. 4 A minimal subset of 26 YAC clones spanning the entire genome of the fission yeast *S. pombe* is shown aligned to available genetic[10] and physical[11] map information. The dotted lines beside of those indicating the genetic markers used are representing cosmids, which were hybridized to *Not*I-digests of *S. pombe* as well.

mapping system for this organism, and provide access to all its DNA, but could also serve as a substrate for a complete sequence determination of the genome. Together with the sequence analyses on human and *S. cerevisiae*[26], both organisms as evolutionarily distant from each other as from *S. pombe*, the detailed analysis of *S. pombe* should provide extremely valuable information for the understanding of their biology.

Methodology

High density gridded YAC library. The YAC library was prepared as described previously[27]. Genomic *S. pombe* DNA (strain 972h$^-$) was cloned in the vector pYAC4[12]; 1248 colonies were picked into 13 microtitre dishes and stored at -70 °C. Six identical copies of the YAC library were gridded directly from microtitre plates onto 22×22 cm nylon filters (Hybond-N$^+$, Amersham). The filters were pre-moistened in synthetic dextrose (SD) medium (2% glucose, 0.7% yeast nitrogen base without amino acids, 1.4% casamino acids, 100 μgml^{-1} adenine hemisulphate and 55 μgml^{-1} tyrosine) containing tryptophan (0.02%) and calcium propionate (0.25%) as an antifungal agent. Gridding was done by a semi-automatic robot system, which comprises a simple X-Y-Z platform and a head designed to accommodate 96-pin gadgets. Colonies were allowed to develop for 2–3 days at 30 °C, on 2% agar of the same composition as the filter-wetting solution.

The colony-lysis procedure has been described previously[28]. In brief, filters are transferred to an absorbent paper pad (Whatman 3MM) soaked in 1 M sorbitol, 0.1 M sodium citrate, 10 mM disodium EDTA (SCE) containing 10 mM dithiothreitol and 4 mgml^{-1} Novozym (Novobiolabs). Cells are converted to spheroplasts by overnight treatment at 37 °C under these conditions. Filters are exposed to denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 15 mins at room temperature and 4 min in the steam of a waterbath of 95 °C, then neutralised on 1.5 M NaCl, 1.0 M Tris-HCl pH 7.2 for 5 min. Filters are then digested with 0.25 mgml^{-1} proteinase K (BDH) in 150 mM NaCl, 100 mM Tris-HCl pH 7.2, for 30–60 min at 37 °C. Filters are dried at room temperature and the DNA crosslinked to the filter with UV light from a germicidal lamp.

Probe isolation and hybridization. Cosmid DNA was prepared by an alkaline lysis procedure[29]. DNA of the genetic markers was obtained by PCR amplification of already known sequences. YAC ends were isolated as described[23]. The YAC filters were prehybridized at 65 °C in 7% sodium dodecyl sulphate (SDS), 0.5 M sodium phosphate (pH 7.2), 1 mM EDTA and 0.1 mgml^{-1} yeast tRNA for 1–2 h. Hybridization of oligolabelled probes[30] was carried out overnight under the same conditions. Subsequently, the filters were rinsed twice at room temperature in 40 mM sodium phosphate (pH 7.2), 0.1% SDS and washed at 65 °C in the same buffer for 10–20 min. Exposure was at -70 °C for 2–3 days using intensifying screen. The hybridization results were digitized and stored in our *S. pombe* database.

Data analysis. Data analysis was carried out on a SUN SPARC-II workstation, using software written in C and Prolog. Our database hybridization data from a variety of sources, for example, optical densities from computer-based image-analysis, x-y-coordinates from a digitiser and text files of clone names. These input formats are translated into a canonical format, listing the positive clones for each experiment and merging the resulting data for each round of analysis. In one ordering method, the distance between each possible pair of probes is calculated from their hybridization frequencies to the YAC clones as the percentage of clones hybridising only to one probe but not to both. The probes are then ordered by finding the shortest circuit connecting all the probes using simulated-annealing[31] forming contiguous arrays of probes which are kept separate whenever the distance between two neighbouring probes exceeds a given value. Another program exploits a set of heuristic rules, that excludes potentially chimaeric clones and repetitive elements from the contig construction, and inferred minimal clone and probe sets covering the chromosomes. A minimum of three linking clones were required to join any pair of probes into a contig so as to prevent false connections. Probes from repetitive regions of the genome were deleted iteratively until each repetitive element was surrounded by two non-repetitive probes which could be linked by at least three clones.

***Not*I-digestion of YAC DNA.** Yeast colonies were grown on SD plates containing tryptophan (0.02%) and calcium propionate (0.25%). DNA was prepared in agarose blocks from the yeast cultures as described[28]. *Not*I digestion was carried out overnight at 37 °C with an enzyme:DNA ratio of about 20:1. Uncut YAC DNA and their *Not*I-digests were electrophoresed on a CHEF apparatus at 160 volt at a 100-s switch time for 27 h. The gel was alkali blotted to a nylon filter (Hybond-N$^+$, Amersham) and hybridized with pBR322, a probe specific to the YAC vector, to verify the sizes of the respective YAC clones.

Received 3 April; accepted 21 May 1992.

Acknowledgements

We thank Mitsuhiro Yanagida and Antony Carr for providing genetic probes of S. pombe, *Paul Nurse for the* S. pombe *strain 972h⁻ and suggestions on the manuscript and Ross Overbeck for discussions about the data analysis. We also thank Janet Roden, Neil Burden, Lisa Gellen, Christal Douglas, Marion Denter and Catherine Knights for their technical help. E. M. is funded by an European Community Research Fellowship.*

1. Kaufer, N. F., Simanis, V. & Nurse, P. *Nature* **318**, 78–80 (1985).
2. Russell, P. & Nurse, P. *Cell* **45**, 781–782 (1986).
3. Allshire, R. C. *et al. Cell* **50**, 391–403 (1987).
4. Lee, M. G. & Nurse, P. *Nature* **327**, 31–35 (1987).
5. Uemura, T. *et al. Cell* **50**, 917–925, (1987).
6. Booher, B. & Beach, D. *Cell* **57**, 1009–1016 (1989).
7. Moreno, S., Hayles, J. & Nurse, P. *Cell* **58**, 361–372 (1989).
8. Kinoshita, N., Ohkura, H. & Yanagida, M. *Cell* **63**, 405–415 (1990).
9. Kohli, J. *Current Genetics* **11**, 575–589 (1987).
10. Munz, P., Wolf, K., Kohli, J. & Leupold, U. *Molecular Biology of the Fission Yeast* (eds Nasim, A. *et al.*) 1–30 (Academic Press, San Diego, 1989).
11. Fan, J. *et al. Nuc. Acid. Res.* **17**, 2801–2818 (1989).
12. Burke, D. T., Carle, G. F. & Olson, M. V. *Science* **236**, 806–812 (1987).
13. Lehrach, H. *et al. In Genome Analysis: Genetic and Physical Mapping* (eds Davies, K. E. & Tilghman, S.) 39–81 (Cold Spring Harbor Laboratory Press, New York, 1990).
14 Olson, M. V., Hood, L., Cantor, C. R. & Botstein, D. *Science* **245**, 1434–1435 (1989).
15. Chikashige, Y. *et al. Cell* **57**, 739–751 (1989).
16. Schaak, J., Mao, J. & Söll, D. *Nuc. Acid. Res.* **10**, 2851–2864 (1982).
17. Levin, H. L., Weaver, D. C. & Boeke, J. D. *Molec. cell. Biol.* **10**, 6791–6798 (1990)
18. Toda, T., Nakaseko, Y., Niwa, O. & Yanagida, M. *Curr. Genet.* **8**, 93–97 (1984).
19. Fan, J., Grothues, D. & Smith, C. L. *Nuc. Acid. Res.* **19**, 6289–6294 (1991).
20. Kohli, J., Hottinger, H., Munz, P., Strauss, A. & Thuriaux, P. *Genetics* **87**, 471–489 (1977).
21. Hirano, T., Funahashi, S., Uemura, T. & Yanagida, M. *EMBO J.* **5**, 2973–2979 (1986).
22. Broughton, B. C. *et al. Molec. gen. Genet.* **228**, 470–472 (1991).
23. Riley, J. *et al. Nuc. Acid. Res.* **18**, 2887–2890 (1990).
24. Little, R. D., Pilia, G., Johnson, S., D'Urso, M. D. & Schlessinger, D. *Proc. natn. Acad. Sci. U.S.A.* **89**, 177–181 (1992).
25. Sternberg, N. L. *Proc. natn. Acad. Sci. U.S.A.* **87**, 103–107 (1990).
26. Oliver, S. G. *et al. Nature* **357**, 38–46 (1992).
27. Larin, Z., Monaco, A. P. & Lehrach, H. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4123–4127 w(1991).
28. Larin, Z. & Lehrach, H. *Genet. Res.* **56**, 203–208 (1990).
29. Birnboim, H. C. & Doly, J. *Nuc. Acid. Res.* **7**, 1513–1523 (1979).
30. Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6–13 (1983).
31. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes in C* 343–351 (Cambridge University Press, Cambridge, 1988).