Volume 15 Number 12 1987 Nucleic Acids Research

A novel sequence common to the centromere regions of *Schizosaccharomyces pombe* chromosomes

Yukinobu Nakaseko*, Noriyuki Kinoshita and Mitsuhiro Yanagida

Department of Biophysics, Faculty of Science, Kyoto University, Sakyo-ku, Kyoto 606, Japan

Received April 24, 1987; Accepted May 22, 1987 Accession no. Y00342

ABSTRACT

An approximately 4 kb long sequence (designated dh) is located in the centromere regions of all three chromosomes of S. pombe. There is one copy each of dh per centromere in chromosomes I and II and multiples in the centromere of chromosome III. Nucleotide sequence determination shows that dhI and dhII are highly homologous. A part of the sequence (ca. 300-400 bp) contains short direct repeats, otherwise dh is in general internally non-repetitious. Although there are three segmental deletions (total 821 bp) and two insertions (27 bp) in dhII (an 80% overall homology to dhI), there are only nine substitutions between dhI and dhII in the remaining 3980 bp, giving a 99.77% homology. The substitutions are restricted to the non-repetitious domains and are only of the pyrimidine-pyrimidine or purine-purine types. A possible conformational role of dh is discussed.

INTRODUCTION

The centromere is a functional domain of chromosomes, the attachment site for kinetochore microtubules which radiate from the centrosomes to separate the chromosomes during mitosis (1-4). The centromere is also the site for sister chromatid association. Prior to the S phase, each chromosome consists of a single sister chromatid and, after chromosomal DNA is replicated in the S phase, two sister chromatids of each chromosome are tightly held together at the centromere. Thus the centromere is an important structure for chromosome separation and maintenance. Molecular basis for microtubule attachment and sister chromatid association is little understood.

In the budding yeast Saccharomyces cerevisiae, the centromere DNAs have been cloned from different chromosomes (5-12) and defined as approximately 200 bp sequences, consisting of three domains. These sequences appear to act as centromeres functionally and structurally, when they are joined to other chromosomal elements such as ARS sequences and telomeres in synthetic minichromosomes and introduced into cells of S. cerevisiae.

The nucleotide sequence organization of the centromere structure in the fission yeast Schizosaccharomyces pombe is quite different (13, 14). There

© IRL Press Limited, Oxford, England.

appears to be a roughly 60 kb long centromere domain, consisting of a number of repeating sequences, in which meiotic recombination is greatly reduced (13). For example, a sequence ynl is repeated three times only in the centromere II region. Another sequence nk is repeated roughly 20 times in the genome, three times in the centromere II region and at least twice in the centromere I, but not in the centromere III region (unpublished result). Among these repeating sequences, one sequence called dg is common to all the centromere regions (13). The dg family comprises the 4 kb long sequences that are highly homologous each other. A common centromeric sequence such as dg should possibly play an important role in the centromere functions of S. pombe. Our recent experiments suggest that dg is essential for maintenance and proper segregation of chromosomes in mitosis and meiosis (Y. Nakaseko et al., to be published). We have examined whether sequences other than dg exists in the centromere regions, and will report here on one such sequence we have named dh. The nature of the nucleotide sequences as well as their localizations in the centromere domains are described.

MATERIALS AND METHODS

Standard procedures for DNA preparation, restriction enzyme digestion, gel transfer and Southern hybridization were followed (13, 15, 16). Filters were hybridized in 5XSSPE and 0.3% SDS using 100 μg/ml salmon sperm DNA. The filters were washed in 10 mM $NaHPO_4$, 1mM EDTA and 0.2% SDS at room temperature. For nucleotide sequence determination, DNA fragments were subcloned into pUC18 and pUC19 (17).

Cosmids used in the present study were described previously (13). YIp5 (18) was a vector for pSS161 (dhI), pSS162 (dhI) and pSS166 (dgIIa) while pUC18 for pSS241 (dhII) and pSS242 (dhII).

RESULTS

Identification of a novel centromeric sequence in S. pombe

To determine whether any sequence other than dg is in common to the centromere I and II regions of S. pombe, a series of Southern blot hybridizations was done, using centromere II cosmids probed with a collection of centromere I-linked sequences cloned into YIp5 or pUC18 (Materials and methods). The three cosmid clones B4, C1 and C2 covering a 100 kb long region which includes the entire centromere II (13) were restricted with Hind III, run on 0.8% agarose gels, blotted and probed with several DNA fragments derived from the centromere I region (Figure 1a).

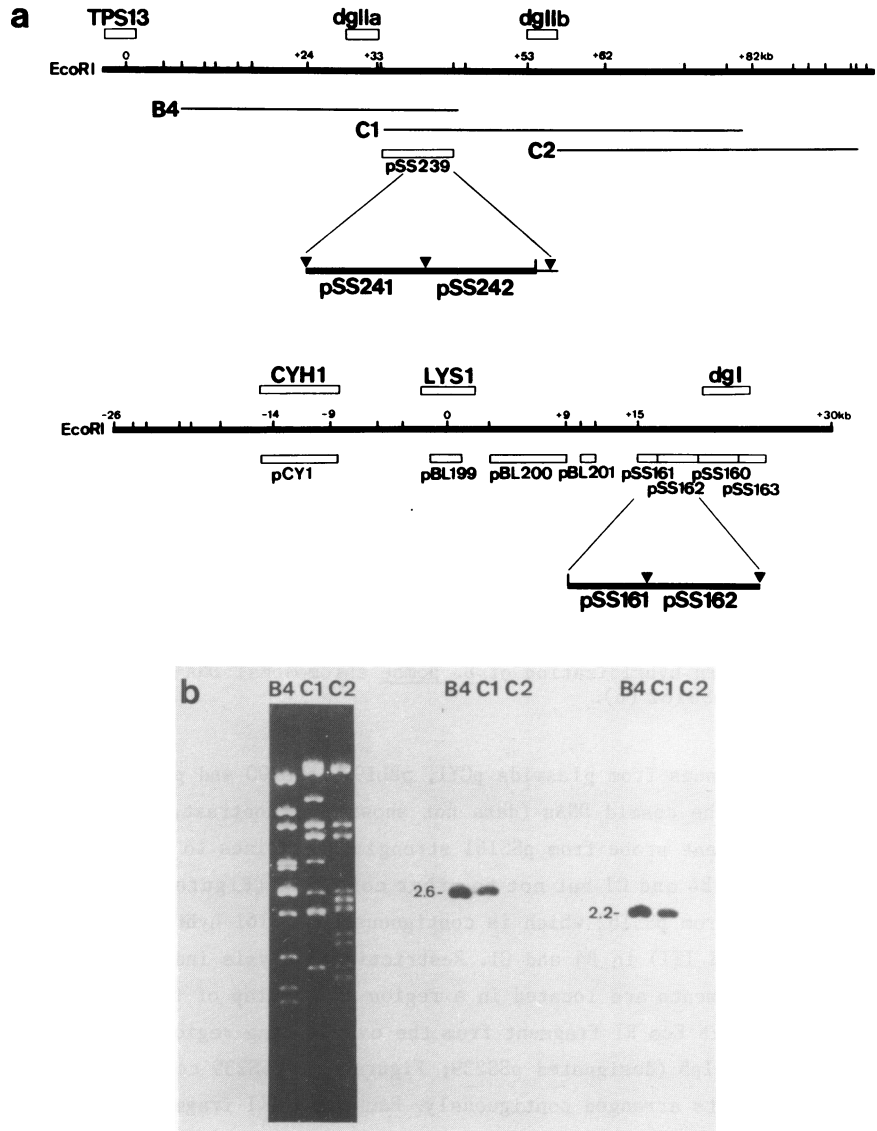

Figure. 1. (a) Markers and Eco RI restriction maps in the vicinity of the centromeres of chromosome I (bottom) and II (top). Three cosmid clones B4, C1 and C2 derived from the centromere II are shown. Filled triangle, Hind III sites. (b) Southern blot hybridization of cosmids B4, C1 and C2 digested with Hind III and probed with pSS161 and pSS162. Ethidium-bromide stained agarose gel electrophoretic patterns digested with various restriction enzymes (left), autoradiogram probed with pSS161 (center) and probed with pSS162 (right).

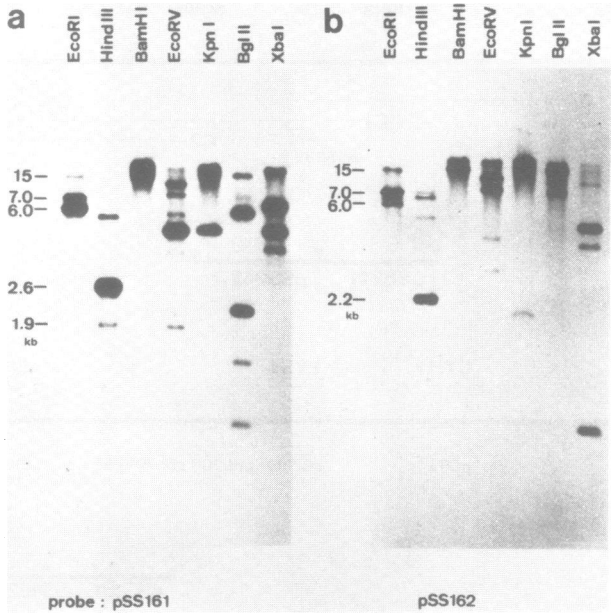

Figure 2. Southern hybridization of S. pombe chromosomal DNA probed with pSS161 (a) and pSS162 (b).

Probe sequences from plasmids pCY1, pBL199, pBL200 and pBL201 did not hybridize with the cosmid DNAs (data not shown). In contrast, a 1.7 kb Hind III-Eco RI fragment probe from pSS161 strongly hybridizes to a 2.6 kb Hind III band in cosmids B4 and C1 but not to other cosmid C2 (Figure 1b). The 2.2 kb Hind III probe from pSS162 which is contiguous to pSS161 hybridizes to another band (2.2 kb Hind III) in B4 and C1. Restriction analysis indicates that these hybridizing fragments are located in a region of overlap of the two cosmids, B4 and C1. An 8 kb Eco RI fragment from the overlapping region in B4 was subcloned into YIp5 (designated pSS239; Figure 1a). pSS239 contains the two Hind III fragments arranged contiguously. Each Hind III fragment was further subcloned into pUC plasmids (pSS241 hybridizable to pSS162 and pSS242 hybridizable to pSS161; Figure 1a). Sequence analysis described below showed that the two fragments contained a novel centromeric sequence (called dh). The other probes pSS160 and pSS163 contain dgI (13).

Genomic Southern hybridization probed with pSS161 and pSS162

S. pombe genomic DNA was isolated, digested with various restriction enzymes, run on 0.8% agarose gels and hybridized with pSS161 or pSS162 under

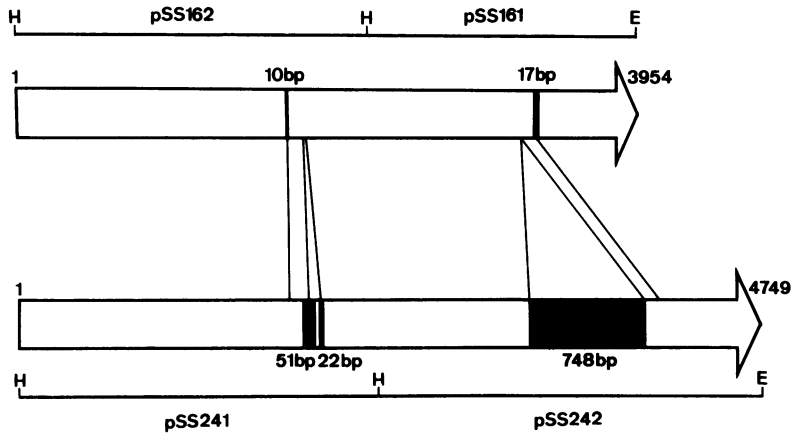

Figure 3. Nucleotide sequence determination of dhI and dhII. Nucleotide sequences of pSS161, pSS162, pSS241 and pSS242 were determined. Deletions and insertions are indicated. E, Eco RI; H, Hind III. The arrows indicate the direction of dh sequence.

stringent conditions (Figure 2). Multiple hybridizing bands are seen, indicating that more than two such sequences are present in the genome. Their copy number although low (probabaly 5-10) has not been exactly determined because the bands appear to overlap and the degree of homology to the probes is not known.

By genomic Southern hybridization, other three probes from pCY1, pBL199 and pBL200 were shown to represent unique genomic sequence, while the probe from pBL201 is repetitive (data not shown).

Nucleotide sequence determination

The dideoxy method was employed to determine a 8703 bp long nucleotide sequence: 1708 bp Eco RI-Hind III in pSS161 and 2246 bp Hind III in pSS162 (3954 bp dhI derived from the centromere I region); 2440 bp Eco RI-Hind III in pSS241 and 2309 bp Hind III in pSS242 (4749 bp dhII derived from the centromere II region). The arrangement and the direction of the determined sequences are shown in Figure 3. The sequences from the two centromere regions are highly homologous, and the homologies extend to the ends of the determined sequences. For comparison of the two nucleotide sequences (Figure 4), the full sequence is shown for dhII with the nucleotide number (1-4749) while only the differences are shown for dhI.

In dhI, there are two segmental insertions (10 bp at 1718 and 17 bp at 3307), three segmental deletions (51 bp at 1830, 22 bp at 1847 and 748 bp at

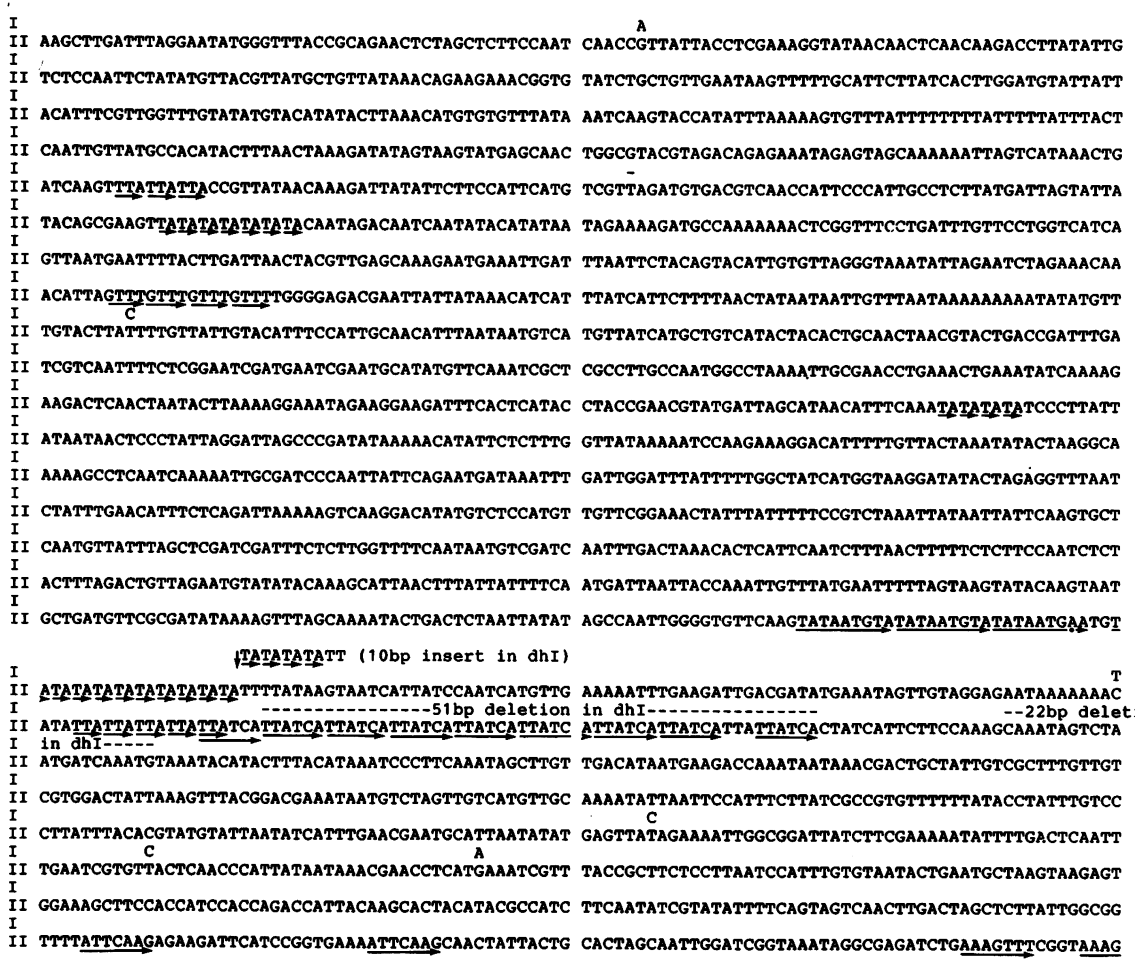
(10bp insert in dhI)
51bp deletion in dhI
22bp deletion

Figure 4. Nucleotide sequences of dhI and dhII. See text.

Table 1 Short repeating units in dh

Repeating unit	No. of repeats dhI	dhII	Total length (bp) dhI	dhII
TTA	8	8	24	24
TA	25	21	50	42
GTTT	4	4	16	16
TATAATGNA	3	3	27	27
TTATCA	0	9	0	54
ATTCAA	2	2	12	12
AAAGTTT	2	2	14	14
TTTC	2	2	8	8
TTTTN (N=G or C)	3	3	15	15
NCATATXYAZMT (N=T,C: X=T,C: Y=C,A: Z=G,T: M=G,T)	4	4	48	48
TCTCTATA	0	4	0	32
TTNXAY (N=A,G: X=T,C: Y=G,A)	0	11	0	66
GNATGGXTATGGACACA (N=C,G: X=A,G)	2	1	34	17
CANG (N=A,T)	4	4	16	16
GNATTA (N=A,T)	3	3	18	18
Total			282	409

3214) and one nucleotide deletion (at 455). Therefore, the total length of dhI determined is 3954 bp, 795 bp shorter than that of dhII. Several classes of short direct repeats (total about 300-400 bp) are scattered (underlined in Figure 4 and summarized in Table 1). The longest repeating unit is 17 bp long, GNATGGNTATGGACACA. A 30-bp long TA stretch is present in dhI. Most of the segmental deletions or the insertions involve stretches of the short direct repeats. Neither inverted repeats nor coding frames have been found.

The overall homology between dhI and dhII is approximately 80%. Nonhomology is mainly due to the large 748 bp deletion in dhI. Substitutions, however, rarely occur; they are scattered and are present only in the non-direct repeat regions. There are only 9 replacements in 3927 bp, giving an extremely high homology (99.77%). Interestingly, substitutions are restricted to the type of purine-purine (G-A) or pyrimidine-pyrimidine (T-C).

Occurrence of the dh sequence in the centromere III region

Two lines of evidence show that the centromere III also contains the dh sequences. (1) pSS161 strongly hybridizes to different fragments in the cosmid clones (AB5, BB1, CB12 and DD7; Figure 5a, b) which contain dgIII and are derived from the centromere III region (ref. 13 and unpublished result). (2) pSS161 strongly hybridizes to the PFG gel electrophoretic band of an S. pombe mini-chromosome Ch16 (19) which contains the centromere III region (data not

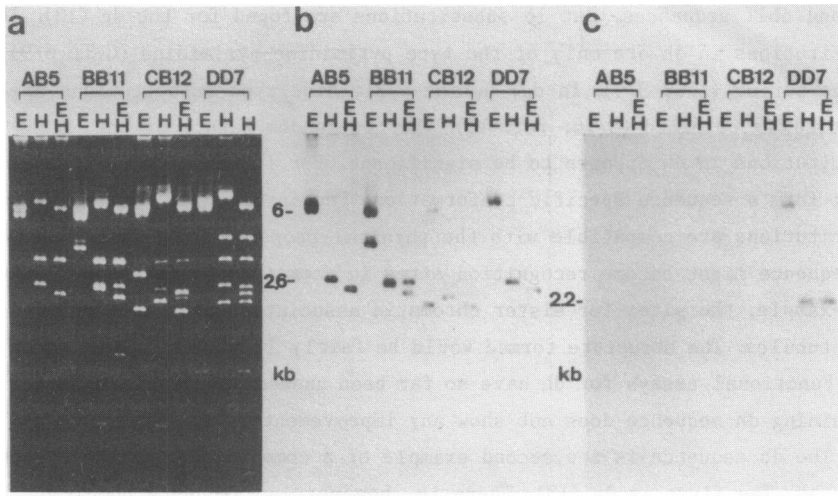

Figure 5. (a) Ethidium bromide stained agarose gel electrophoretic patterns of four cosmids AB5, BB11, CB12 and DD7 derived from the centromere III, digested with Eco RI (E), Hind III (H) and both Eco RI and Hind III (EH) and run in agarose gel. Blot hybridizations of the cosmids probed with pSS161 (b) and pSS162 (c).

shown). Only one cosmid DD7 hybridizes to pSS162 (Figure 5c), indicating that a part of dhI (and dhII) is missing in certain dhIII sequences. The dhIII sequences are not all identical. Because walking has not been completed in the centromere III, precise locations of these cosmids in the centromere III is not known.

Discussion

We report in this paper that an 4-5 kb long highly homologous sequence that we have named dh is common to the three centromere domains of S. pombe chromosomes. Computer analysis indicated that dh does not appear to encode for any protein. Although it contains stretches of short direct repeats, most of the sequence is not internally repetitious, unlike highly repetitious satellite DNAs in the centromere regions of higher eucaryotes. There is only one dhII sequence in centromere II. One dhI is found in centromere I, but because walking is not completed in this region, more than one dhI might be present. In the centromere III, multiple dhIII sequences appear to be present.

The role of the dh sequences is unknown. The high degree of homology between dhI and dhII suggests that these sequences are involved in an important centromere function. Only 9 substitutions are found in the 4-kb long

dhI and dhII sequences. But 16 substitutions are found for the dg (13). The substitutions in dh are only of the type pyrimidine-pyrimidine (C-T: 6/9) or purine-purine (A-G: 3/9). In dg, purine-pyrimidine type substitutions are plentifully obtained (5/16; ref. 13). The non-random nature of the substitutions in dh appears to be significant. For instance, the dh sequences might form a sequence specific conformation. Perhaps only A-G or C-T substitutions are compatible with the physical property to be maintained. The dh sequence might become recognition sites in common to different centromeres, for example, the sites for sister chromatid association and kinetochore microtubules. The structure formed would be fairly large, containing 4 kb of DNA. Functional assays for dh have so far been unsuccessful; plasmids containing dh sequence does not show any improvement in stability.

The dh sequence is the second example of a common centromere sequence in S. pombe. The first is dg (13). There is, however, no sequence homology between dg and dh. Not all of the dh sequences in the genome might be necessarily restricted to the centromere regions. The exact copy number of dh is not determined; it is probably in the range of 5-10. All of the dg sequences, on the other hand, appear to be present in the centromeres (13). The copy number of dg in the genome is 7-8. The dh sequences tend to be linked to dg: dhI and dhII are adjacent to dgI and dgIIa, respectively. However, dg is not always associated with dh: dgIIb has no dh in its neighbourhood (Fig. 1). The distance between dhI and dgI is about 1 kb, while that between dhII and dgIIa is about 3 kb (unpublished result). The junction sequences between dg and dh remain to be determined.

Acknowledgements

We thank Dr. John Pulitzer for reading the manuscript. This work was supported by grants from the Mitsubishi Foundation, the Ministry of Education, Science and Culture and the Science and Technology Agency of Japan.

*Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

REFERENCES

1. Blackburn,E.H. and Szostak,J.W. (1984) Ann. Rev. Biochem. 53, 163-194.
2. Murray,A.W. and Szostak,J.W. (1985) Ann. Rev. Cell Biol. 1, 289-315.
3. Ris,H and Witt,P.L. (1981) Chromosoma 82 153-170.
4. Pickett-Heaps,J.D., Tippit,D.H. and Porter,K.R. (1982) Cell 29, 729-744.
5. Clark,L. and Carbon,J. (1980) Nature 287, 504-509.
6. Hsiao,C.-L. and Carbon,J. (1981) Proc. Natl. Acad. Sci. USA 78, 3760-3764.
7. Stinchcomb,D.T., Mann,C. and Davis,R.W. (1982) J. Mol. Biol. 158, 157-179.

8. Fitzgerald-Hays,M., Buhler,J,M. and Cooper,T.G. and Carbon,J. (1982) Mol. Cell Biol. 2, 82-87.
9. Panzeri,L and Phillipsen,P. (1982) EMBO J. 1,1605-1611.
10. Maine,G.S., Surosky,R.T. and Tye,B.-K. (1983) Mol. Cell Biol. 4, 86-91.
11. Hieter,P., Pridmore,D., Hegamann,D.H., Thomas,M., Davis,R.W. and Philippsen,P. (1985) Cell 42, 913-921.
12. Clarke,L. and Carbon,J. (1985) Annu. Rev. Genet. 19, 29-56.
13. Nakaseko,Y., Adachi,Y., Funahashi,S., Niwa,O. and Yanagida,M. (1986) EMBO J. 5, 1011-1021.
14. Clarke,L., Amstutz,H., Fishel,B. and Carbon,J. (1986) Proc. Natl. Acad. Sci. USA 83, 8253-8257.
15. Maniatis,T., Fritsch,E.A. and Sambrook,J. (1982) Molecular Cloning Cold Spring Harbor Laboratory, New York.
16. Botstein,D., Falco,S.C., Stewart,S.E., Brennan,M., Schara,S., Stinchcomb,D.T., Struhl,K. and Davis,R.W. (1979) Gene 8, 17-24.
17. Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) Gene 33, 103-119.
18. Struhl,K., Stinchcomb D.T. and Davis,R.W. (1979) Proc. Natl. Acad. Sci. USA. 76, 1035-1039.
19. Niwa,O., Matsumoto,T., Yanagida,M. (1986) Mol. Gen. Genet. 203, 397-405.