

# A Family of Telomere-Associated Autonomously Replicating Sequences and Their Functions in Targeted Recombination in *Hansenula polymorpha* DL-1

JUNG-HOON SOHN,<sup>1</sup> EUI-SUNG CHOI,<sup>1</sup> HYUN AH KANG,<sup>1</sup> JOON-SHICK RHEE,<sup>2</sup>  
AND SANG-KI RHEE<sup>1\*</sup>

Biotechnology Research Division, Korea Research Institute of Bioscience and Biotechnology,  
Yusong, Taejeon 305-600,<sup>1</sup> and Department of Biological Sciences, Korea Advanced  
Institute of Science and Technology, Yusong, Taejeon 305-701,<sup>2</sup> Korea

Received 29 June 1998/Accepted 23 November 1998

**A family of multiple autonomously replicating sequences (ARSs) which are located at several chromosomal ends of *Hansenula polymorpha* DL-1 has been identified and characterized. Genomic Southern blotting with an ARS, HARS36, originating from the end of a chromosome, as a probe showed several homologues in the genome of *H. polymorpha*. Nucleotide sequences of the three fragments obtained by a selective cloning for chromosomal ends were nearly identical to that of HARS36. All three fragments harbored an ARS motif and ended with 18 to 23 identical repetitions of 5'-GGGTGGCG-3' which resemble the telomeric repeat sequence in other eukaryotes. Transformation of *H. polymorpha* with nonlinearized plasmids containing the newly obtained telomeric ARSs almost exclusively resulted in the targeted integration of a single copy or multiple tandem copies of the plasmid into the chromosomes. The sensitivity to exonuclease *Bal31* digestion of the common DNA fragment in all integrants confirmed the telomeric origin of HARS36 homologues, suggesting that several chromosomal ends, if not all of them, consisted of the same ARS motif and highly conserved sequences observed in HARS36. Even though the frequencies of targeted recombination were varied among the ends of the chromosomes, the overall frequency was over 96%. The results suggested that the integration of the plasmids containing telomeric ARSs occurred largely through homologous recombination at the telomeric repeats, which serve as high-frequency recombination targets.**

The methylotrophic yeast *Hansenula polymorpha* has attracted much attention as a host for the production of recombinant proteins (for reviews, see references 12 and 20). One of the reasons might be its unusual property of genetic recombination between a transforming plasmid and the chromosome. The multiple tandem integration of up to 100 copies of a nonlinearized plasmid has been observed in the chromosome via homologous and nonhomologous recombination (2, 23, 35). Several unusual observations related to the recombination, such as plasmid reorganization by the insertion of chromosomal DNA (4, 16) and the amplification of a gene copy number after integrative transformation, have also been reported (14). No report, to date, has been made about the natural plasmid in *H. polymorpha*, and as found in other yeasts, autonomously replicating sequences (ARSs) of *H. polymorpha* that can act as the replication origin have been obtained either from its chromosome (4, 35, 40) or other sources (3, 42). The frequency of transformation was significantly increased when an ARS element was employed. A transforming plasmid, even one with an ARS, showed a very low mitotic stability of less than 5% for 10 generations, but cells with extremely high mitotic stability could easily be obtained after some dozens of generations in a nonselective condition, resulting from the integration of a plasmid into the chromosome (35, 40). During this step, interestingly, multiple tandem copies of a plasmid are often integrated into the chromosome. Due to the lack of intensive studies of *H. polymorpha* at the molecular level, it is

uncertain whether an episomal plasmid with an ARS can be maintained consistently, due to its higher frequency of recombination. Furthermore, the mechanism of multiple tandem integration is also unknown.

From the biotechnological point of view, this characteristic of *H. polymorpha* provides a valuable tool for the development of industrial strains producing recombinant proteins. Stable maintenance of high copy numbers of gene expression cassettes could greatly improve the productivity of recombinant proteins in long-term fermentation. Unfortunately, however, the frequency of obtaining multiple integrants, up to 100 copies as described above, appears to be quite low and unpredictable. We previously suggested the importance of the ARS employed to the increased frequency of multiple tandem integration (40). An ARS of *Hansenula*, HARS36, was obtained from the genome of *H. polymorpha* DL-1 by an enrichment procedure to produce an ARS with characteristics of tandem repeat integration. HARS36 increased the frequency of transformation and multiple tandem integration of plasmids into chromosomes. In a functional analysis of HARS36, three important domains for the episomal replication and integration into chromosomes were identified as regions A, B, and C. Especially, region C of HARS36 was responsible for the enhanced rate of chromosomal integration (40). Transformation of a plasmid containing HARS36 and a selection marker facilitated the selection of an integrant with over 100 tandem copies of a plasmid in chromosomes (2). Furthermore, the integration copy number was elaborately controlled by the use of a dominant selection marker (39). Strikingly, most of the integration events with a plasmid containing HARS36 appeared to occur near different ends of chromosomes, suggesting that the origin of HARS36 related to the ends of chromosomes (39,

\* Corresponding author. Mailing address: Biotechnology Research Division, Korea Research Institute of Bioscience and Biotechnology, Yusong, P.O. Box 115, Taejeon 305-600, Korea. Phone: 82-42-860-4005. Fax: 82-42-860-4592. E-mail: rheesk@kribb4680.kribb.re.kr.

40). Here, we demonstrated that HARS36 is a family of telomeric ARSs residing in several ends of chromosomes in *H. polymorpha*. Region C of HARS36, which enhanced chromosomal integration of a plasmid, was found to be a telomeric repeat of *H. polymorpha* that has not been identified to date. Additionally, we also demonstrated that the ends, if not all, of *H. polymorpha* chromosomes consisted of sequences similar to HARS36 and served as high-frequency recombination targets for a plasmid containing HARS36.

#### MATERIALS AND METHODS

**Strains, media, and plasmids.** *H. polymorpha* DL1-L (*leu2*) derived from DL-1 (ATCC 26012) (26) was kindly provided by M. Y. Beburow (Moscow, Russia). Yeast cells were grown on yeast extract-peptone-dextrose (YPD) medium (1% [wt/vol] yeast extract, 2% [wt/vol] Bacto Peptone, and 2% [wt/vol] glucose). Transformants obtained with a plasmid containing *LEU2* were selected in a minimal selective medium (0.67% [wt/vol] yeast nitrogen base without amino acid and 2% [wt/vol] glucose). *Escherichia coli* DH5 $\alpha$  [F<sup>-</sup>lacZ $\Delta$ M15 *hsdR17* (r<sup>-</sup>m<sup>-</sup>) *gyrA36*] was used for the general recombinant DNA techniques. The plasmid pCE36 containing HARS36 and *LEU2* of *H. polymorpha* (1, 40) was used as a control plasmid for the measurement of the transformation frequency and the DNA source of HARS36. Plasmids pGEM-CE120 and pGE-MARS286 containing deletion derivatives of HARS36 (40) were used to obtain the DNA fragments for labeling of hybridization probes. Plasmids pBluescript KS+ and pBC KS+ (Stratagene, La Jolla, Calif.) were used for the general DNA cloning and sequencing. Plasmids pBKS-TEL188, -135, and -61 contained one of three telomeric fragments that were obtained in this study. Plasmids pCTEL188, -135, and -61 were derived from pCHLX containing *H. polymorpha* *LEU2* (40) after subcloning of the *HindIII*-*Bam*HI fragments of pBKS-TEL188, -135, and -61, respectively.

**General DNA techniques.** General DNA manipulations were performed as described by Sambrook et al. (37). DNA fragments required for subcloning and labeling experiments were gel purified with a QiaEx kit (Qiagen, Valencia, Calif.). Total yeast DNA was isolated with Novozym (Novo Biolabs, Bagsvaerd, Denmark) for large-scale preparation and with glass beads for small-scale preparation, as described by Johnston (24). Plasmid DNA isolation from *E. coli* was done by the boiling method described in the work of Sambrook et al. (37). Nucleotide sequencing was carried out by dideoxy sequencing reaction with an ABI Prism dye terminator cycle sequencing ready reaction kit (Perkin-Elmer Cetus, Foster City, Calif.), and the sequences were read in an automatic DNA sequencer (model 373A; Applied Biosystems, Foster City, Calif.). *E. coli* was transformed by the Simple and Efficient Method (22).

**Transformation of *H. polymorpha* and stabilization of the transformant.** The transformation of *H. polymorpha* was performed according to the lithium acetate-dimethyl sulfoxide method described by Hill et al. (19). Initial transformants showing heterogeneous colony sizes due to different mitotic stabilities were stabilized with a procedure to introduce a transforming plasmid into the chromosome. For the stabilization, all transformants selected for *LEU2* prototrophy in minimal selective medium were pooled with YPD broth and inoculated into a 500-ml baffled flask containing 50 ml of YPD broth medium. After 24 h, 1 ml of culture broth was transferred to 50 ml of fresh YPD medium. This procedure was repeated until the cells had reached 50 generations. Then, the cells were plated on minimal selective plates to select for *LEU2* prototrophy again. Complete stabilization was monitored by spreading cells onto minimal selective plates and checking for an even growth rate, judging from the sizes of colonies. After 100% stability in the *LEU2* phenotype of the individual stabilized colony was confirmed by the replica plating of cells on minimal and complex media, the colonies were stored under selective conditions.

**Cloning of telomeric fragments.** For the cloning of telomeric fragments, total genomic DNA was first treated with T4 DNA polymerase to fill in the 3' overhangs of chromosome ends, prior to digestion with *Pst*I. Treatment with T4 DNA polymerase was done in a mixture of 50 mM Tris-HCl (pH 8.5), 5 mM MgCl<sub>2</sub>, 5 mM dithiothreitol, 50  $\mu$ g of bovine serum albumin per ml, and 0.1 mM concentrations of deoxynucleoside triphosphates in a 50- $\mu$ l volume for 1 h at room temperature. After digestion with *Pst*I for 4 h at 37°C, the DNA sample was fractionated by 1% agarose gel electrophoresis. DNA fragments of 300 to 600 bp isolated from the agarose gel were ligated to the plasmid pBluescript KS+, which was linearized with *Pst*I and *Eco*RV to make libraries of *E. coli*. Among transformants, 50 to 100 subsets covering 0.5  $\times$  10<sup>3</sup> to 1  $\times$  10<sup>3</sup> colonies in total were tested for hybridization with the P3 probe (Fig. 1). After the screening of subsets, individual clones were isolated from positive subsets by the same procedure.

**Bal31 digestion.** Thirty micrograms of genomic DNA was digested at 30°C with 1 U of *Bal*31 nuclease (New England Biolabs, Beverly, Mass.) in a total volume of 500  $\mu$ l of 600 mM NaCl–20 mM Tris-HCl (pH 8.0)–12 mM CaCl<sub>2</sub>–12 mM MgCl<sub>2</sub>–1 mM EDTA. A 3- $\mu$ g sample was recovered before the addition of the enzyme and at several time intervals after the enzyme addition. Ten microliters of 0.25 M EGTA was added to stop the reaction, and DNA was pelleted with 2 volumes of ethanol. This DNA was used for further digestion with restriction endonuclease.

**Southern hybridization.** Genomic Southern hybridization was done as described by Sambrook et al. (37). Total chromosomal DNA was isolated and digested with restriction endonucleases. After electrophoresis, DNA was capillary transferred onto a nylon membrane (Schleicher & Schuell GmbH, Dassel, Germany). The labeling of probe DNA was performed with a non-radioactive DNA labeling and detection kit (Boehringer Mannheim, Mannheim, Germany). Digoxigenin labeling of probes P1 and P3 was carried out with ARS fragments from pCE36 and pGE-MARS286 as templates, respectively. Probes P2 and P4 were labeled with the *Ava*I-*Kpn*I fragment of pCE36 and the *Hind*III fragment of pGEM-CE120, respectively. Probes pBC and *LEU2* were made from pBC KS+ (Stratagene) and the *Bam*HI-*Xba*I fragment of pCHLX containing *H. polymorpha* *LEU2* (1), respectively. All hybridization was carried out at 42°C in a hybridization oven (Hybaid, Middlesex, United Kingdom) with a hybridization solution (5 $\times$  SSC [1 $\times$  SSC is 0.15 M NaCl plus 0.015 M sodium citrate], 0.1% [wt/vol] *N*-lauroylsarcosine, 0.02% [wt/vol] sodium dodecyl sulfate, 5% [wt/vol] blocking reagent, and 50% [vol/vol] formamide) as recommended in the manufacturer's instructions.

**Nucleotide sequence accession numbers.** The nucleotide sequences of HARS36, TEL188, TEL135, and TEL61 have been submitted to the GenBank database under the accession numbers U31858, U82170, U82171, and U82172, respectively.

#### RESULTS

**Genomic origin of HARS36.** Previously, we reported on an analysis of several functional elements of an ARS of *Hansenula*, HARS36 (Fig. 1A), which has the ability to introduce multiple copies of a plasmid containing this sequence into the chromosomal end (40). HARS36 has three important functional elements: regions A and B are required for episomal replication, and region C is responsible for the high tendency of integration into a chromosome. The persistent telomeric localization of the transforming plasmids with HARS36 strongly suggested that HARS36 may also originate from the end of a chromosome (40). To get some information on the genomic origin of HARS36, we carried out nucleotide sequencing of region C and found that region C contains 18 copies of a highly regular 8-bp G-rich repeating unit (5'-GG GTGGCG-3'). The repeating unit resembled the telomeric repeat sequence of several other yeasts and other eukaryotic organisms, as shown in Table 1, supporting the presumption that HARS36 originated from the end of a chromosome.

For the further analysis of the chromosomal origin of HARS36, genomic Southern blots were carried out with HARS36 as a probe (Fig. 1B). Surprisingly, more than 10 bands were detected, even though the signal intensities varied among the bands. Sequence homologues to HARS36 appeared to exist in multiple copies in the genome, as in the case of multiple ARS families found near the ends of chromosomes in *Saccharomyces cerevisiae* (6). Telomeric and subtelomeric regions of *S. cerevisiae* are composed of conserved sequences such as Y' and X sequences which contain ARSs (7, 46). Thus, it is conceivable that HARS36 is a family of ARSs that reside near the ends of chromosomes in *H. polymorpha*. To determine which part of HARS36 is responsible for the multiplicity, further genomic Southern blot analyses were done with three probes, P2, P3, and P4, from different parts of HARS36 (Fig. 1A). Probe P2, containing both ARS parts (regions B and A) and the repeated sequence (region C) of HARS36, and probe P3, containing only the ARS part, were used. Probe P4 was labeled from region D of HARS36, which has no effect on ARS activity after deletion (40). As shown in Fig. 1B, P2 also hybridized more than 10 chromosomal DNA bands, which is almost the same number as P1. Interestingly, the P3 probe, which has no repeat sequence (region C), hybridized with only five signals strongly detected with P1 and P2. Thus, weak bands detected with P2 appeared to be detected by the repeat sequences in region C of HARS36. In contrast to these, probe P4 hybridized with a unique signal that was also detected with the P1 probe. These results suggested that the multiplicity of

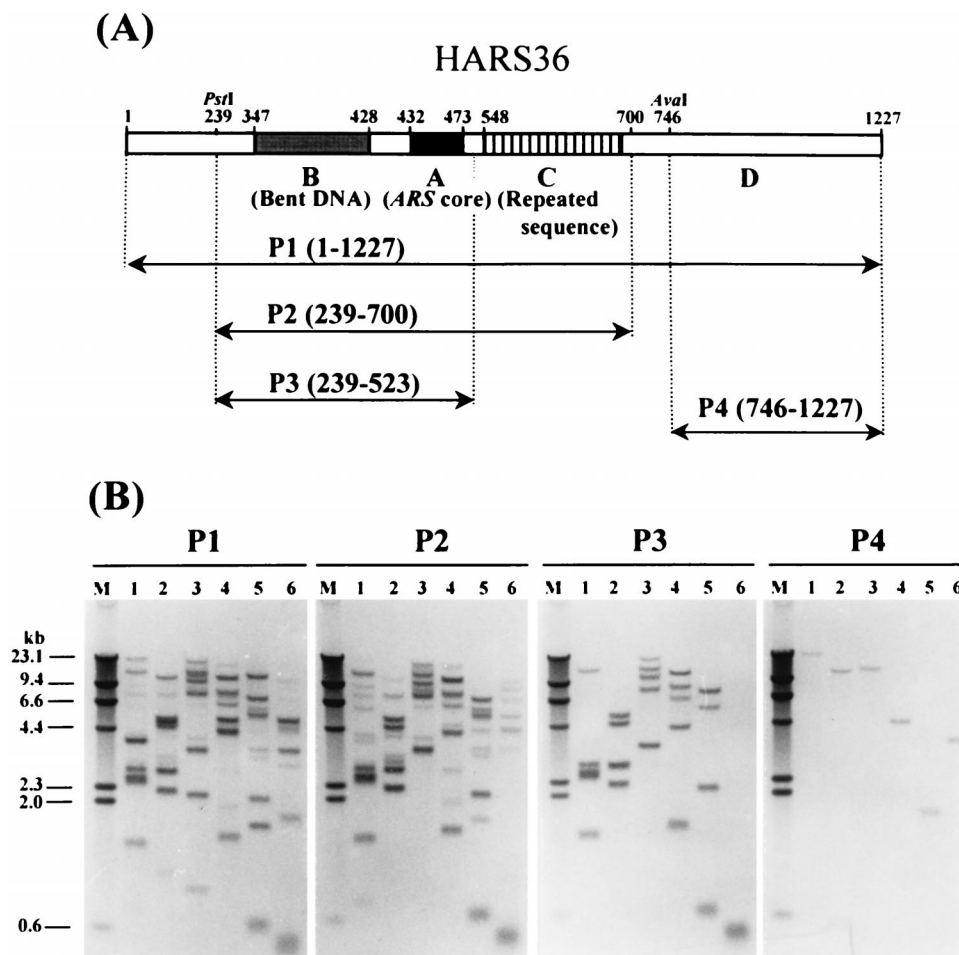


FIG. 1. Physical map of HARS36 (A) and genomic Southern blots (B). Each of the functional elements A, B, and C was described in Sohn et al. (40). Four different probes, P1 to P4, were digoxigenin-labeled from different parts of HARS36, as shown on the map. Probe P1 was labeled from a DNA fragment containing the entire (base pairs 1 to 1227) HARS36, P2 from nucleotide 239 to 700, P3 from nucleotide 239 to 523, and P4 from nucleotide 746 to 1227. The hybridization was performed as described in Materials and Methods. Lane M, digoxigenin-labeled lambda *Hind*III size marker (23.1, 9.4, 6.6, 4.4, 2.3, 2.0, and 0.6 kb); in other lanes, genomic DNA digested with *Eco*RI (lane 1), with *Hind*III (lane 2), with *Xba*I (lane 3), with *Xho*I (lane 4), with *Bam*HI (lane 5), and with *Pst*I (lane 6).

HARS36 in the genome was largely determined by the sequence between nucleotides 239 and 700 of HARS36 containing the whole ARS motif (40). Several homologues to the ARS motif containing regions A, B, and C of HARS36 appeared to be present in the chromosome of *H. polymorpha*. Furthermore, several additional homologues to region C of HARS36 but without regions A and B also appeared in the chromosome.

**Cloning of a family of multiple telomeric ARSs.** From the results of the telomeric localization of a transforming plasmid containing HARS36 (40) and the resemblance of the repeat sequence of region C to other telomeric repeats, we assumed that the chromosomal fragments of *H. polymorpha* hybridized with P2 might contain an identical telomeric repeat sequence, but only five among them harbored a common ARS element (regions A and B). Depending on the restriction enzymes used, P3 hybridized with one to five chromosomal fragments (Fig. 1B). Only a single band of about 450 bp was detected when genomic DNA was digested with *Pst*I, suggesting that all five telomeric fragments detected with P3 have a highly conserved sequence including a *Pst*I site near the end of the chromosome. To check this, *Pst*I-digested genomic DNA fragments of 300 to 600 bp, which were hybridized with the probe P3, were cloned into *Pst*I-digested pBluescript KS+. About 10<sup>3</sup> colonies were screened with probe P3, but we failed to get positive clones.

If all five *Pst*I fragments of 450 bp represent the ends of different chromosomes, an end of each *Pst*I fragment would be the physical end of a chromosome. Because chromosomes often end with a 3' overhang of about 10 nucleotides (46), the

TABLE 1. Comparison of the repeat sequence of region C in HARS36 to telomeric repeats of yeasts and other organisms (18)

Organism	Sequence
<i>H. polymorpha</i>	GGGTGGCG
<b>Yeasts</b>	
<i>S. cerevisiae</i>	T(G) <sub>2-3</sub> (TG) <sub>1-6</sub>
<i>S. pombe</i>	TTAC(A) <sub>2-5</sub>
<i>C. guilliermondii</i>	ACTGGTGT
<i>K. lactis</i>	ACGGATTTGATTAGGTATGT GGTGT
<b>Others</b>	
<i>Tetrahymena</i>	TTGGGG
<i>Oxytricha</i>	TTTTGGGG
<i>Arabidopsis</i>	TTAGGG
<i>Homo sapiens</i>	TTAGGG



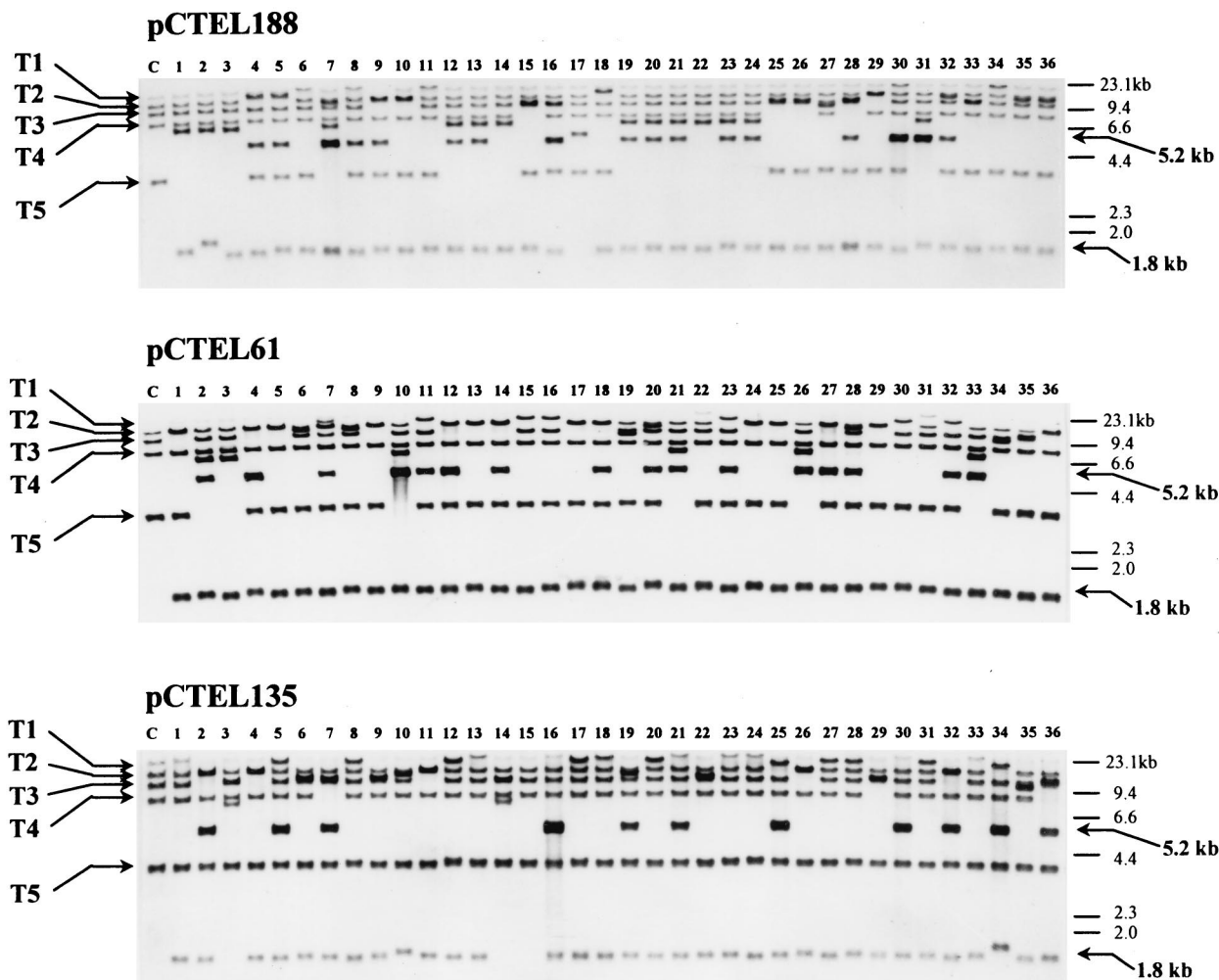


FIG. 3. Integration patterns of transformants with plasmids pCTEL188, -61, and -135. Three sets of 36 randomly selected integrants were obtained by transformation with three plasmids, pCTEL188, -61, and -135, containing a telomeric ARS. All genomic DNAs were digested with *Xba*I and probed with P3 (Fig. 1). T1 to T5 are the five P3 homologues of the *Xba*I-digested genomic DNA of untransformed *H. polymorpha* DL-1. The 1.8- and 5.2-kb bands are, respectively, a common band that appeared when integration occurred near the end of chromosomes and a band that appeared when integration occurred in a tandem repeated array of a plasmid.

were transformed into *H. polymorpha*. Each pool of transformants was completely stabilized to integrate a transforming plasmid into the chromosomes, as described in Materials and Methods. After stabilization, 36 individual colonies of each pool grown on minimal selective plates were randomly selected to check their integration patterns. Total genomic DNA was isolated from individual colonies and digested with *Xba*I. Southern blotting was done with a P3 probe that could hybridize with the five *Xba*I-digested genomic fragments in five P3 homologues (Fig. 1). Southern blots of three sets of 36 integrants with the P3 probe are shown in Fig. 3. Interestingly, most integrants showed different patterns of five P3 homologues (T1 to T5) with those found in untransformed cells (Fig. 3, lane C). It indicated that the patterns of five P3 homologues were modified by the integration of a transforming plasmid. Judging from the change of Southern blot patterns in five P3 homologues, we concluded that a transforming plasmid containing a telomeric ARS integrated into one of the five P3 homologues with a considerably high frequency (72 of 108 integrants tested).

If all five P3 homologues were different end fragments of chromosomes, the possible patterns of integration into five P3

homologues would be as it is depicted in Fig. 4. In the case of a single-copy integration into one of five P3 homologues that is located in the physical ends of the chromosomes, two modifications might be expected in the *Xba*I-digested band patterns of P3 homologues. One modification could be the size upshift of about 3.4 kb in a band corresponding to the homologue, and the other could be the appearance of a new band (1.8 kb) corresponding to the fragment containing *H. polymorpha* *LEU2* and the telomeric ARS. The size of the new *Xba*I band should be the same in all integrants into P3 homologues, because an end of the band is the physical end of the chromosome. As expected, a new band of about 1.8 kb is common to all 72 integrants showing a size shift of a band corresponding to one of the five P3 homologues (Fig. 3). Surprisingly, the 1.8-kb band is also detected in the integrants without a size shift of P3 homologues (32 of the remaining 36 integrants). This fact suggested that the integration also occurred in end fragments other than the five P3 homologues. Such end fragments were not detected with P3 but were weakly detected with P2 containing the telomeric repeated sequence (Fig. 1B). Small size variations of this 1.8-kb fragment in the integrants appeared to be caused by the differences in positions of integration or in

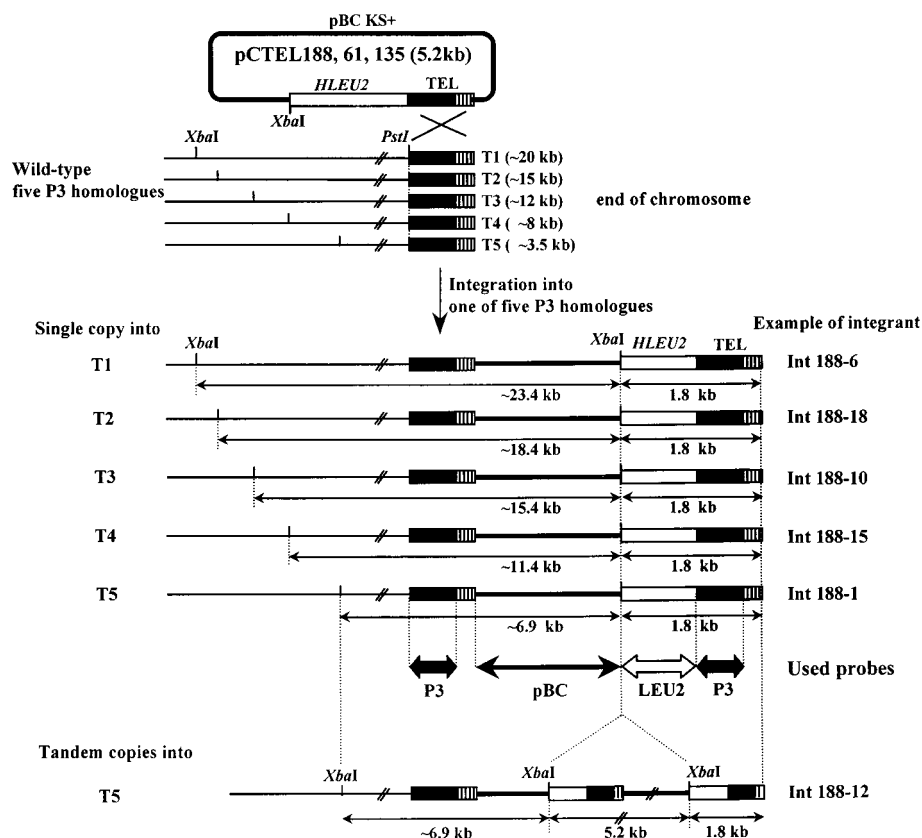


FIG. 4. Schematic diagram of telomeric integrations into five P3 homologues (T1 to T5). The sizes of five P3 homologues were roughly estimated from the Southern blot results shown in Fig. 3. Black boxes and striped boxes indicate that the telomeric conserved region up to the *Pst*I site containing the ARS and telomeric repeats were conserved in five different ends of chromosomes. Empty boxes and bold lines represent the *H. polymorpha* *LEU2* gene and pBC KS+, respectively. Probes used in this study are shown under the corresponding regions as two-headed arrows. An example of each integrant showing a single-copy integration of pCTEL188 into one of the five P3 homologues was selected from the Southern blot results shown in Fig. 3. Numbers after Int188 to the right are lane numbers.

numbers of the telomeric repeat. In the case of tandem integration of a transforming plasmid into a single locus, an additional 5.2-kb signal would be generated, which corresponds to the internal copies of the transforming plasmid located inside the repeat of plasmids (Fig. 4). The 5.2-kb band was detected in 44 integrants, in which a 1.8-kb band also appeared (Fig. 3). These facts strongly suggested that a single or multiple integration of the plasmids containing newly cloned ARSs were exclusively targeted to the different ends of chromosomes via homologous recombination.

**Telomeric location of five P3 homologues.** To confirm the telomeric integration of the plasmids and the telomeric origins of five P3 homologues, five different integrants of pCTEL188 (Int188-6, Int188-18, Int188-10, Int188-15, and Int188-1), which harbored a single-copy integration into one of five P3 homologues (T1 to T5), were selected and further characterized. As shown in Fig. 5, genomic Southern blots were done with three different probes, P3, pBC, and LEU2. All five different integrants showed a size-shifted P3 homologue that was not detected in untransformed cells (Fig. 5A). Each size-shifted P3 homologue could also be detected with probe pBC, indicating that the size shift of each P3 homologue was caused by the integration of a transforming plasmid (Fig. 5B). A common new signal of about 1.8 kb was detected with probes LEU2 and P3 but not with pBC (Fig. 5C). A genomic single copy of *H. polymorpha* *LEU2* (6.5 kb) was also detected with probe LEU2. To prove the terminal localization of the *Xba*I-

digested 1.8-kb band, possibly a new chromosomal end fragment, the sensitivity of the 1.8-kb signal to *Bal*31 was tested. Genomic DNA of Int188-1 was treated in time course with endonuclease *Bal*31 and then digested with *Xba*I. Southern blotting of the DNA was done with LEU2 as a probe (Fig. 6). The internal endogenous *H. polymorpha* *LEU2* signal of 6.5 kb was not sensitive to *Bal*31 digestion. In contrast to this, the size of the 1.8-kb signal was progressively shortened by incubation with *Bal*31. It indicated that the 1.8-kb fragment common to all integrants (except 4 of 108) is a new end of a chromosome. It also indicated that all five *Xba*I-digested P3 homologues are different ends of chromosomes.

**Frequency of targeted recombination.** The pattern and frequency of each plasmid's integration into the different chromosomal ends are presented in Table 2 by an analysis of the Southern blot shown in Fig. 3. Though the transforming plasmids integrated largely into the telomeric regions of chromosomes whose sequences are highly homologous to P3 (T1 to T5), they also appeared to integrate into the other telomeric fragments that have less homology to P3. The number of chromosomes of *H. polymorpha* was reported to be six, based on the number of DNA bands resolved by pulsed-field gel electrophoresis (13). Thus, besides five P3 homologues, seven additional telomeric fragments should occur as a group of less homology to P3. All plasmids containing one of the three telomeric ARSs integrated into the telomeric locus with an overall frequency of 96.3%, indicating that the integration oc-

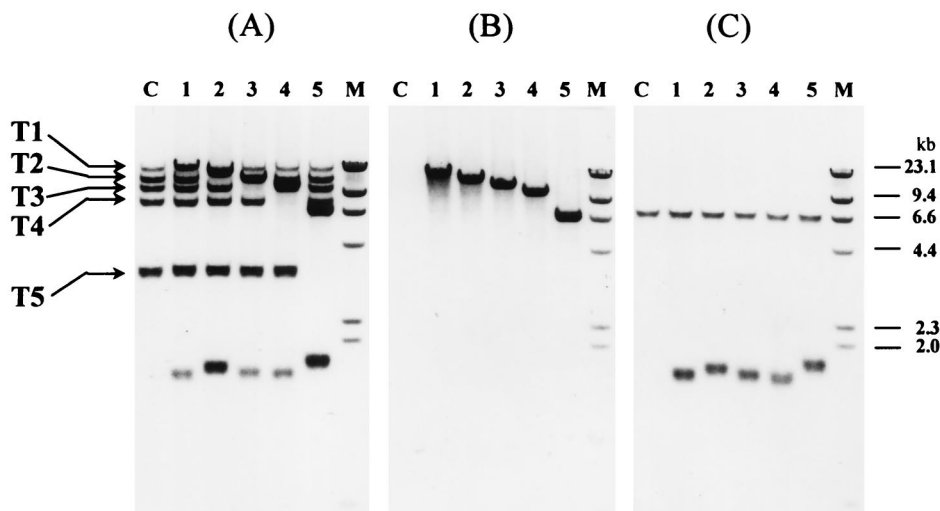


FIG. 5. Confirmation of integrations of pCTEL188 into the five P3 homologues (T1 to T5) with different probes. Five single-copy integrants of pCTEL188 (lanes: 1, Int188-6; 2, Int188-18; 3, Int188-10; 4, Int188-15; and 5, Int188-1) were selected from the results shown in Fig. 3 and analyzed with three probes, P3 (A), pBC (B), and LEU2 (C) (Fig. 4). All genomic DNA was digested with *Xba*I. (C, untransformed *H. polymorpha*; M, digoxigenin-labeled lambda *Hind*III size marker [23.1, 9.4, 6.6, 4.4, 2.3, 2.0, and 0.6 kb]).

curred almost exclusively via homologous recombination. The telomeric repeats that exist in all chromosomal ends might be partly ascribed to this high frequency of homologous recombination responsibility. In addition, two telomeric ARSs, TEL188 and TEL61, introduced the plasmid into P3 homologues (T1 to T5) with much higher frequencies than TEL135. The frequency of targeted integration of TEL188 and TEL61 into P3 homologues was around 75%, but that of TEL135 was around 47%. The striking difference between these two subtypes is the long deletion of a part of the bent region (region B) of the ARS in TEL135. Thus, it is conceivable that the bent regions present in TEL188 and TEL61 play a role in the targeted integration into P3 homologues. Higher frequencies of tandem integration with TEL188 (47.2%) and TEL61 (44.4%) than that with TEL135 (30.6%) (Table 2) also appeared to be related to this sequence.

DISCUSSION

The Southern blot analysis with HARS36 as a probe suggested that five telomeres of *H. polymorpha* chromosomes appear to contain sequences highly homologous to HARS36. We have isolated and characterized three telomeric fragments among the HARS36 homologues that harbored a common ARS motif and telomeric repeats. In a functional study, a derivative of HARS36 which lacks the telomeric repeats, region C, was more effective than the entire HARS36 for prolonged autonomous replication in an episomal state (40). The combination of an ARS domain and a repeated sequence domain greatly increased the potential of HARS36 for the multiple integration of plasmids containing these domains. It was

TABLE 2. Comparison of frequencies of integration into telomeres for telomeric ARSs

Integration locus	Frequency of integration (%) <sup>a</sup>		
	TEL188	TEL61	TEL135
<b>P3 homologues</b>			
T1	8.3	5.6	22.1
T2	8.3	16.7	5.6
T3	8.3	38.9	13.9
T4	11.1		8.3
T5	36.1	16.7	
<b>Other telomeres</b>			
	25.1	22.1	44.5
<b>No telomeres</b>			
	2.8		8.3
<b>Total</b>	<b>100 (47.2)<sup>b</sup></b>	<b>100 (44.4)</b>	<b>100 (30.6)</b>

<sup>a</sup> The telomeric integration frequencies of three telomeric ARSs were measured by counting the number of each telomeric integration from the Southern blot results shown in Fig. 3.

<sup>b</sup> Figure in the parentheses is the integration frequency of the tandem repeated array of a plasmid.

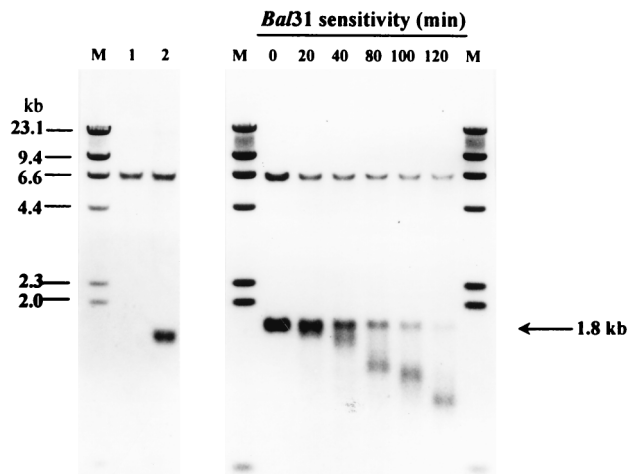


FIG. 6. *Bal*31 sensitivity of the common 1.8-kb band. Genomic DNA of Int188-1, which harbored a single integration of pCTEL188 into P3 homologue T5, was treated with exonuclease *Bal*31 for various amounts of time. Subsequently, DNA was digested with *Xba*I. Southern blotting was done with probe LEU2. (lane 1, untransformed *H. polymorpha* without *Bal*31 treatment; lane 2, Int188-1 without *Bal*31 treatment; M, digoxigenin-labeled lambda *Hind*III size marker [23.1, 9.4, 6.6, 4.4, 2.3, 2.0, and 0.6 kb]).

noted that the integration events always occurred near the ends of different chromosomes, suggesting that a highly conserved sequence environment may also exist near the ends of different chromosomes in *H. polymorpha*, as found in other eukaryotic organisms (7, 9, 32, 36).

Telomeric regions of yeasts have been known to be highly recombinogenic. There has been evidence of frequent recombination between telomeres (21) and between plasmids and chromosomes (10). Targeted recombination in this study suggested that two high-frequency recombination targets exist, telomeric repeats (5'-GGGTGGCG-3') and a bent region in the ARS motif, in the ends of chromosomes of *H. polymorpha*. It has been observed that telomeric repeats serve as a recombinational hot spot (5, 25). Repressor activator protein (RAP1) of *S. cerevisiae* is an essential nuclear protein that recognizes a 13-bp consensus sequence found in numerous upstream activating sequences at the silencers of transcriptionally repressed mating-type genes and in telomeric repeats (38). Interestingly, RAP1 was found to stimulate the telomeric repeat-mediated recombination (15). Binding of RAP1 to a site of the *HIS4* gene also stimulates recombination at this locus (34, 45), suggesting that the high frequency of targeted recombination in this study could be mediated by the similar protein(s) of *H. polymorpha*. Investigation of the telomeric binding protein(s) could provide insight into the nature of recombination in *H. polymorpha*. Involvement of intrinsically bent DNA elements in the enhanced recombination has also been reported for other organisms (11, 29). Though the exact biological function of the bent structure in the DNA is still unclear, it has been speculated that it modifies the chromatin structure of DNA and thus facilitates the binding of related protein factors for DNA replication, transcription, and recombination (17).

In the case of TEL135, which has a large deletion near the bent region, a decrease in frequencies of both tandem array integration and targeted integration into the five P3 homologues was observed, compared to the frequencies for TEL61 and TEL188. TEL135 also showed decreased ARS activity. This observation indicated that the bent structure of the subtelomeric region in *H. polymorpha* might play a role in replication or recombination, both of which might affect the frequency of the tandem integration of plasmids. Two mechanisms of plasmid integration in tandem array could be conceivable: either the integration of plasmid in a single copy number and subsequent integration of next copies of plasmid into the same site via homologous recombination or the integration of preformed dimeric (or multimeric) plasmid into the chromosome. Based on this, we suggest that the lower frequency of tandem array integration for TEL135 could be the consequence of either a lower recombination efficiency or a lower proportion and/or copy number of autonomous multimeric plasmid before stabilization. Further intensive studies are required to elucidate the mechanism that produces the tandem repeated arrays in *H. polymorpha*.

The present study revealed that the chromosomes of *H. polymorpha* ended with identical repeats, an 8-bp G-rich sequence (5'-GGGTGGCG-3'), as found in other eukaryotic chromosomes (18). Generally, telomeric repeated sequences consist of a simple sequence in tandem repeats of 5 to 9 bp and with a G-rich strand running 5'→3' towards the end of the chromosome. Telomeric repeats from several other yeasts, such as *S. cerevisiae* (44), *Schizosaccharomyces pombe* (27), several *Candida* species, and *Kluyveromyces lactis* (28), have been reported. The telomeric repeats (5'-GGGTGGCG-3') of *H. polymorpha* started from 80 bp downstream of the putative ARS core of HARS36. This repeating unit was extended by 18 to 23

repetitions to make up about 144 to 184 bp in total. It showed an unusually high G+C content of over 87% and a G content of over 75%. It appeared to be the highest G+C content among those of telomeric repeats reported to date (18). The G+C content of telomeric repeats is often inversely related to the overall G+C content of the genomes (33). Such an extremely high G+C content for the telomeric repeats appeared to be exceptional for *H. polymorpha*, for which the genomic G+C content was 48% (31). Another interesting feature of the telomeric repeat of *H. polymorpha* is the highly regular pattern of repeats compared with those of conventional yeasts, such as *S. cerevisiae* [T(G)<sub>2-3</sub>(TG)<sub>1-6</sub>] and *S. pombe* [TTAC(A)G<sub>2-5</sub>]. Several *Candida* species and *K. lactis* showed the regular and longer repeating units (up to 26 bp), except for the 8-bp regular repeat (ACTGGTGT) found in *Candida guilliermondii* (28). Recently, another 8-bp regular repeat (TCTGGGTG) was also found in different yeasts, such as *Saccharomyces castellii* and *Saccharomyces dairensis* (8).

In addition to the simple repeats found at the very ends of chromosomes, subtelomeric regions of chromosomes in many organisms often contain repetitive elements called telomere-associated sequences (18). Very little is known about the origin or function of these elements (46). Structures of telomeres have been studied in detail in *S. cerevisiae* (6, 7, 41). Most of the chromosomal ends of *S. cerevisiae* were found to be associated with two conserved elements, Y' and X. A repetitive ARS family and tandem repeats of a short G-rich sequence [T(G)<sub>2-3</sub>(TG)<sub>1-6</sub>] were also found to be associated with the Y' and X elements (7, 43). It is apparent from our results that the subtelomeric region of several chromosomal ends of *H. polymorpha* also contained such conserved elements containing a family of the ARS motif. All single-copy integrants of a transforming plasmid into different chromosomal ends in this study might be valuable for telomere analysis. Plasmid rescues into *E. coli* after religation of *Xba*I-digested genomic fragments from single-copy integrants could make it possible to obtain the flanking fragments containing subtelomeric regions from different chromosomes. Information about individual chromosomes with a telomeric structure will provide us with a tool for the genome mapping of *H. polymorpha*, which has been little investigated.

#### ACKNOWLEDGMENT

This work has been funded by the Ministry of Science and Technology of Korea. We deeply appreciate its financial support.

#### REFERENCES

1. Agaphonov, M. O., A. I. Poznyakovski, A. I. Bogdanova, and M. D. Ter-Avanesyan. 1994. Isolation and characterization of the *LEU2* gene of *Hansenula polymorpha*. *Yeast* 10:509-513.
2. Agaphonov, M. O., P. M. Trushkina, J.-H. Sohn, E.-S. Choi, S.-K. Rhee, and M. D. Ter-Avanesyan. *Yeast*, in press.
3. Berardi, E., and D. Y. Thomas. 1990. An effective transformation method for *Hansenula polymorpha*. *Curr. Genet.* 18:169-170.
4. Bogdanova, A. I., M. O. Agaphonov, and M. D. Ter-Avanesyan. 1995. Plasmid reorganization during integrative transformation in *Hansenula polymorpha*. *Yeast* 11:343-353.
5. Brahmachari, S. K., G. Meera, P. S. Sarkar, P. Balagurumoorthy, J. Tripathi, S. Raghavan, U. Shaligram, and S. Pataskar. 1995. Simple repetitive sequences in the genome: structure and functional significance. *Electrophoresis* 16:1705-1714.
6. Chan, C. S., and B. K. Tye. 1983. A family of *Saccharomyces cerevisiae* repetitive autonomously replicating sequences that have very similar genomic environments. *J. Mol. Biol.* 168:505-523.
7. Chan, C. S., and B. K. Tye. 1983. Organization of DNA sequences and replication origins at yeast telomeres. *Cell* 33:563-573.
8. Cohn, M., M. J. McEachern, and E. H. Blackburn. 1998. Telomeric sequence diversity within the genus *Saccharomyces*. *Curr. Genet.* 33:396-400.
9. Corcoran, L. M., J. K. Thompson, D. Walliker, and D. J. Kemp. 1988. Homologous recombination within subtelomeric repeat sequences generates



- chromosome size polymorphisms in *P. falciparum*. *Cell* **53**:807–813.
10. **Dunn, B., P. Szauter, M. L. Pardue, and J. W. Szostak.** 1984. Transfer of yeast telomeres to linear plasmids by recombination. *Cell* **39**:191–201.
  11. **Economides, A. N., D. Everdeen, and N. Panayotatos.** 1996. A shared, non-canonical DNA conformation detected at DNA/protein contact sites and bent DNA in the absence of supercoiling or cognate protein binding. *J. Biol. Chem.* **271**:24836–24841.
  12. **Gellissen, G., C. P. Hollenberg, and Z. A. Janowicz.** 1995. Gene expression in methylotrophic yeasts. *Bioprocess Technol.* **22**:195–239.
  13. **Gellissen, G., C. P. Hollenberg, and Z. A. Janowicz.** 1994. Gene expression in methylotrophic yeasts, p. 195–239. *In* A. Smith (ed.), *Gene expression in recombinant microorganisms*. Marcel Dekker, New York, N.Y.
  14. **Gilbert, S. C., H. van Urk, A. J. Greenfield, M. J. McAvoy, K. A. Denton, D. Coghlan, G. D. Jones, and D. J. Mead.** 1994. Increase in copy number of an integrated vector during continuous culture of *Hansenula polymorpha* expressing functional human haemoglobin. *Yeast* **10**:1569–1580.
  15. **Gilson, E., T. Muller, J. Sogo, T. Laroche, and S. M. Gasser.** 1994. RAP1 stimulates single- to double-strand association of yeast telomeric DNA: implications for telomere-telomere interactions. *Nucleic Acids Res.* **22**:5310–5320.
  16. **Graupner, S., and W. Wackernagel.** 1996. Identification of multiple plasmids released from recombinant genomes of *Hansenula polymorpha* by transformation of *Escherichia coli*. *Appl. Environ. Microbiol.* **62**:1839–1841.
  17. **Hagerman, P. J.** 1990. Sequence-directed curvature of DNA. *Annu. Rev. Biochem.* **59**:755–781.
  18. **Henderson, E.** 1995. Telomere DNA structure, p. 11–34. *In* E. H. Blackburn and C. W. Greider (ed.), *Telomeres*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
  19. **Hill, J., K. A. G. Donald, and D. E. Griffiths.** 1991. DMSO-enhanced whole cell yeast transformation. *Nucleic Acid Res.* **19**:5791.
  20. **Hollenberg, C. P., and G. Gellissen.** 1997. Production of recombinant proteins by methylotrophic yeasts. *Curr. Opin. Biotechnol.* **8**:554–560.
  21. **Horowitz, H., P. Thorburn, and J. E. Haber.** 1984. Rearrangements of highly polymorphic regions near telomeres of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4**:2509–2517.
  22. **Inoue, H., H. Nojima, and H. Okayama.** 1990. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**:23–28.
  23. **Janowicz, Z. A., K. Melber, A. Merckelbach, E. Jacobs, N. Harford, M. Comberbach, and C. P. Hollenberg.** 1991. Simultaneous expression of the S and L surface antigens of hepatitis B, and formation of mixed particles in the methylotrophic yeast, *Hansenula polymorpha*. *Yeast* **7**:431–443.
  24. **Johnston, J. R.** 1988. Yeast genetics, molecular aspects, p. 107. *In* I. Campbell and J. H. Duffus (ed.), *Yeast, a practical approach*. IRL Press, Oxford, England.
  25. **Katinka, M. D., and F. M. Bourgain.** 1992. Interstitial telomeres are hotspots for illegitimate recombination with DNA molecules injected into the macronucleus of *Paramecium primaurelia*. *EMBO J.* **11**:725–732.
  26. **Levine, D. W., and C. L. Cooney.** 1973. Isolation and characterization of a thermotolerant methanol-utilizing yeast. *Appl. Microbiol.* **26**:982–990.
  27. **Matsumoto, T., K. Fukui, O. Niwa, N. Sugawara, J. W. Szostak, and M. Yanagida.** 1987. Identification of healed terminal DNA fragments in linear minichromosomes of *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* **7**:4424–4430.
  28. **McEachern, M. J., and E. H. Blackburn.** 1994. A conserved sequence motif within the exceptionally diverse telomeric sequences of budding yeasts. *Proc. Natl. Acad. Sci. USA* **91**:3453–3457.
  29. **Milot, E., A. Belmaaza, J. C. Wallenberg, N. Gusew, W. E. Bradley, and P. Chartrand.** 1992. Chromosomal illegitimate recombination in mammalian cells is associated with intrinsically bent DNA elements. *EMBO J.* **11**:5063–5070.
  30. **Morrison, A., and A. Sugino.** 1992. Nucleotide sequence of the *POL3* gene encoding DNA polymerase III (delta) of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **20**:375.
  31. **Nakase, T., and K. Komagata.** 1971. Further investigation on the DNA base composition of the genus *Hansenula*. *J. Gen. Appl. Microbiol.* **17**:77–84.
  32. **Pardue, M.-L.** 1995. *Drosophila* telomeres: another way to end it all, p. 339–370. *In* E. H. Blackburn and C. W. Greider (ed.), *Telomeres*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
  33. **Petracek, M. E., P. A. Lefebvre, C. D. Silflow, and J. Berman.** 1990. *Chlamydomonas* telomere sequences are A+T-rich but contain three consecutive G-C base pairs. *Proc. Natl. Acad. Sci. USA* **87**:8222–8226.
  34. **Porter, S. E., M. A. White, and T. D. Petes.** 1993. Genetic evidence that the meiotic recombination hotspot at the *HIS4* locus of *Saccharomyces cerevisiae* does not represent a site for a symmetrically processed double-strand break. *Genetics* **134**:5–19.
  35. **Roggenkamp, R. O., H. Hansen, M. Eckart, Z. A. Janowicz, and C. P. Hollenberg.** 1986. Transformation of the methylotrophic yeast *Hansenula polymorpha* by autonomous replication and integration vectors. *Mol. Gen. Genet.* **202**:302–308.
  36. **Saiga, H., and J. E. Edstrom.** 1985. Long tandem arrays of complex repeat units in *Chironomus* telomeres. *EMBO J.* **4**:799–804.
  37. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
  38. **Shore, D.** 1994. RAP1: a protean regulator in yeast. *Trends Genet.* **10**:408–412.
  39. **Sohn, J.-H., E.-S. Choi, H. A. Kang, J.-S. Rhee, M. O. Agaphonov, M. D. Ter-Avanesyan, and S.-K. Rhee.** Submitted for publication.
  40. **Sohn, J.-H., E.-S. Choi, C.-H. Kim, M. O. Agaphonov, M. D. Ter-Avanesyan, J.-S. Rhee, and S.-K. Rhee.** 1996. A novel autonomously replicating sequence (ARS) for multiple integration in the yeast *Hansenula polymorpha* DL-1. *J. Bacteriol.* **178**:4420–4428.
  41. **Szostak, J. W., and E. H. Blackburn.** 1982. Cloning yeast telomeres on linear plasmid vectors. *Cell* **29**:245–255.
  42. **Tikhomirova, L. P., R. N. Ikononova, and E. N. Kuznetsova.** 1986. Evidence for autonomous replication and stabilization of recombinant plasmids in the transformants of yeast *Hansenula polymorpha*. *Curr. Genet.* **10**:741–747.
  43. **Umek, R. M., M. H. Linskens, D. Kowalski, and J. A. Huberman.** 1989. New beginnings in studies of eukaryotic DNA replication origins. *Biochim. Biophys. Acta* **1007**:1–14.
  44. **Walmley, R. W., C. S. Chan, B. K. Tye, and T. D. Petes.** 1984. Unusual DNA sequences associated with the ends of yeast chromosomes. *Nature* **310**:157–160.
  45. **White, M. A., M. Dominska, and T. D. Petes.** 1993. Transcription factors are required for the meiotic recombination hotspot at the *HIS4* locus in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **90**:6621–6625.
  46. **Zakian, V. A.** 1989. Structure and function of telomeres. *Annu. Rev. Genet.* **23**:579–604.