Nucleic Acids Research, Vol. 20, No. 15 ***4039–4046***

Characterization of *Arabidopsis thaliana* telomeres isolated in yeast

Eric J.Richards*, Steven Chao+, Aurawan Vongs and Jin Yang
Cold Spring Harbor Laboratory, PO Box 100, 1 Bungtown Road, Cold Spring Harbor, NY 11724-2212, USA

Received April 6, 1992; Revised and Accepted July 2, 1992 EMBL accession nos Z12169, Z12170

ABSTRACT

In an effort to learn more about the genomic organization of chromosomal termini in plants we employed a functional complementation strategy to isolate *Arabidopsis thaliana* telomeres in the yeast, *Saccharomyces cerevisiae*. Eight yeast episomes carrying *A. thaliana* telomeric sequences were obtained. The plant sequences carried on two episomes, YpAtT1 and YpAtT7, were characterized in detail. The telomeric origins of YpAtT1 and YpAtT7 insert DNAs were confirmed by demonstrating that corresponding genomic sequences are preferentially degraded during exonucleolytic digestion. The isolated telomeric restriction fragments contain G-rich repeat arrays characteristic of *A. thaliana* telomeres, as well as subterminal telomere-associated sequences (TASs). DNA sequence analysis revealed the presence of variant telomeric repeats at the centromere-proximal border of the terminal block of telomere repeats. The TAS flanking the telomeric G-rich repeat in YpAtT7 corresponds to a repetitive element present at other *A. thaliana* telomeres, while more proximal sequences are unique to one telomere. The YpAtT1 TAS is unique in the Landsberg strain of *A. thaliana* from which the clone originated; however, the Landsberg TAS cross-hybridizes weakly to a second telomere in the strain Columbia. Restriction analysis with cytosine methylation-sensitive endonucleases indicated that both TASs are highly methylated in the genome.

INTRODUCTION

The genomic organization and genetic behavior of the DNA at chromosomal termini display several unusual features. Chromosomal termini in most, possibly all, eukaryotic organisms are capped by tandem arrays of short G-rich repeats (1). The G-rich repeats appear to be necessary and sufficient to ensure the stability and complete DNA replication of the chromosomal termini. Although the primary sequence of the repeats varies between the eukaryotes, there is a rough consensus: $[T/A_{1-4}G_{1-8}]_n$ (2). Within a particular species, the telomeric repeat can be degenerate (*e.g.*, $[TG_{1-3}]_n$ in *S. cerevisiae*), or conform to a strict formula, as does the *Tetrahymena* [TTGGGG] telomere repeat. The presence of two domains within human telomeric repeat blocks was recently demonstrated; the termini of human chromosomes are comprised of uniform $(TTAGGG)_n$ arrays but the centromere-proximal border of the Grich repeat block also contains variant repeats such as TTGGGG and TGAGGG (3).

Study of the DNA sequences which lie adjacent to the telomeric arrays, termed telomere-associated sequences (TASs), has led to further surprises. In the few organisms where TASs have been cloned and studied directly, the telomere-associated regions generally contain moderately reiterated DNA sequences which are frequently found at more than one telomere, and occasionally, at non-telomeric sites (*e.g.*, 4, 5, 6, 7, 8, 9). Intriguingly, the distribution of TASs among particular chromosomes can vary between strains or individuals within a species highlighting the relative fluidity of telomeric regions (5, 9, 10). The variable distribution may result from recombination between non-homologous chromosome ends or reflect frequent rearrangements of TASs (11, 12, 13). Genes embedded within TASs (*e.g.*, *MAL*, *PHO* and *SUC* genes in *S. cerevisiae* (14, 15, 16), VSG genes in trypanosomes (17)) can be altered and/or dispersed throughout the genome by such exchanges and rearrangements. In addition, gene conversion events mediated by TASs have been shown to allow a linear plasmid introduced into *S. cerevisiae* to acquire a functional telomere (*i.e.*, telomere 'healing'), complete with telomeric repeat arrays (18). These and other data indicate that TASs may play a role in chromosome function by providing a 'buffer zone' between the chromosome arm and the terminus in cases of telomere loss (19, 20).

It should be noted that highly reiterated DNA families have been localized by *in situ* hybridization to the telomeric regions, particularly heterochromatic chromosome termini, in many species (*e.g.*, 21, 22, 23). Given the low resolution of *in situ* hybridization techniques, however, it is unclear how close these repeats approach the telomeric repeat arrays and whether they can be properly referred to as TASs.

* To whom correspondence should be addressed

+ Present address: Department of Genetics, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA

In this report, we extend our previous characterization of the telomeric regions of the flowering plant, *Arabidopsis thaliana* (24). In our initial report, we described the isolation of an *A. thaliana* telomeric repeat array in a bacterial plasmid. This telomeric clone demonstrated the conservation of telomere structure in eukaryotes and allowed us to determine the basic structure of *A. thaliana* telomeres: variable length tracts (up to 2.5 kb) of [TTTAGGG] (with G-strand oriented 5′→3′ towards the terminus). We now describe the structure of two *A. thaliana* clones containing both telomeric repeat arrays and flanking telomere-associated sequences isolated by functional complementation using a telomere-deficient yeast artificial chromosome vector.

MATERIALS AND METHODS

Strains, Plants, Plasmids and Probes

E. coli strains DH5α™ (BRL) and DH10B™ (BRL) were used as bacterial hosts, while BF305-15d (a *S. cerevisiae* strain carrying a *ura3* mutation; Bruce Futcher, personal communication) served as the yeast host. Two *A. thaliana* lines were used in this study: the Landsberg ecotype carrying the erecta mutation and the wild type Columbia ecotype. YCF4 (25) and pBluescript SK− (Stratagene) were used as cloning vectors. Hybridization probes were generated from pARR17 (a 3.7 kb *A. thaliana* rDNA subclone; Eric Richards, unpublished) and pAtT4.1 (a derivative of the *A. thaliana* telomeric clone, pAtT4, carrying 0.4 kb of TTTAGGG repeats (24)), while poly (d[CA/TG]) was purchased from Boehringer Mannheim Biochemicals.

Enzymes

Restriction endonucleases were purchased from four sources: New England Biolabs, Boehringer Mannheim Biochemicals, Stratagene and Brisco, Ltd. Nuclease *Bal*31 and T4 DNA ligase were purchased from New England Biolabs. DNA polymerase I Klenow fragment was obtained from Boehringer Mannheim Biochemicals.

Construction and Screening of the *A. thaliana* telomere YAC library

Approximately 15 μg of Landsberg erecta total genomic DNA was cleaved with *Xba*I, followed by incubation with T4 DNA ligase and appropriate buffer (under dilute DNA conditions: 15 μg/4 ml) to circularize nontelomeric restriction fragments (26). Free *Xba*I ends were then partially filled-in with DNA polymerase I Klenow fragment and dCTP + dTTP. The 'one-armed' yeast centromere plasmid, YCF4, (carrying a *URA3* marker and only one functional telomere, in this case a telomere healing substrate, the yeast TAS Y′) was prepared by digestion with *Hin*dIII and *Bgl*II, followed by a partial fill-in reaction using the Klenow fragment and dATP + dGTP. The partially filled-in *Xba*I sites on the genomic fragments were then ligated to the compatible partially filled-in *Hin*dIII sites of the vector. The resulting ligation products were introduced into spheroplasts of the yeast strain, BF305-15d, in the presence of Lipofectin (BRL) (27). Transformants, selected for uracil prototrophy, were screened by colony hybridization with a radiolabeled *A. thaliana* telomere probe. Eight transformants harboring episomes with *A. thaliana* telomeric sequences were found; two of these transformants, YT1 = BF30515d(YpAtT1) and YT7 = BF305−15d(YpAtT7), were characterized.

Plasmid Constructions

To aid subsequent cloning manipulations, YpAtT1 and YpAtT7 were converted into circular derivatives that could be maintained in *E. coli*. To construct the circular derivatives the DNAs were first digested briefly with nuclease *Bal*31 to remove approximately 50 bp from the telomeric ends to create blunt ends. The DNAs were then cleaved at an *Xba*I site in the Y′ portion of the vector to remove the yeast telomere. The *Xba*I site was chosen because we expected the inserts to be devoid of *Xba*I sites since they were derived from *A. thaliana* genomic *Xba*I fragments with modified *Xba*I ends. The YpAtT1 insert, however, had two *Xba*I sites due to the fact that this insert originated from intermolecular ligation of distinct *Xba*I genomic fragments (data not shown). The chimeric nature of the YpAtT1 insert necessitated a partial *Xba*I digestion to leave the insert intact. The *Xba*I ends were converted to blunt ends by Klenow reactions in the presence of all four dNTPs. The DNAs were then circularized under dilute DNA concentrations using T4 DNA ligase, and introduced into an *E. coli* host by electroporation. The circular derivative of YpAtT1, pAtT31, carried the chimeric insert of the parental linear plasmid. The 3.2 kb *Xba*I-*Sac*I fragment carrying the telomeric block and the TAS was removed from pAtT31 and subcloned into pBluescript SK− to produce pAtT32. The sequencing template pAtT33 (See Figure 8) was created by *Pst*I digestion of pAtT32 followed by recircularizing with T4 DNA ligase; these manipulations deleted a 2 kb *Pst*I fragment corresponding to the centromere-proximal portion of the TAS (one *Pst*I site was provided by the vector polylinker). The circular derivative of YpAtT7 was designated pAtT80, and a number of subclones of this telomeric insert were produced using pBluescript SK− and standard methods to facilitate DNA sequencing.

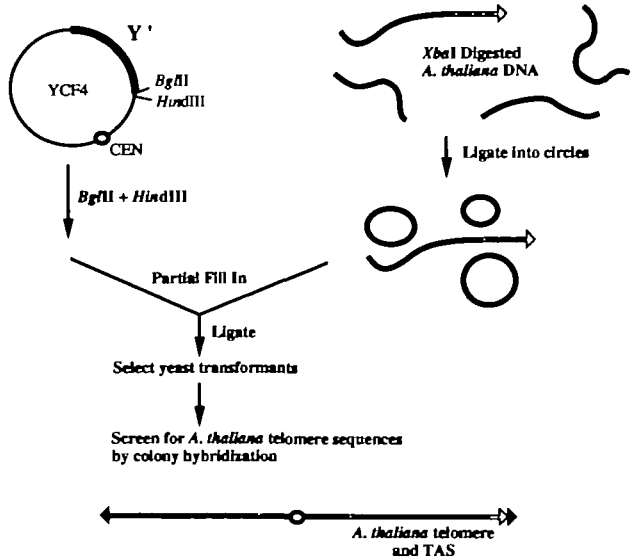

Figure 1. Cloning strategy for isolation of *A. thaliana* telomeres in yeast. The cloning vector, YCF4, is a yeast *URA3* centromere plasmid carrying a single functional telomere. Non-telomeric *A. thaliana* restriction fragments were first depleted by circularizing *Xba*I genomic fragments in a dilute ligation reaction. Telomeric fragments, unable to circularize, were then ligated on to the telomere-less arm of YCF4. The ligation products were introduced into a *ura3* host and uracil prototrophs selected. Transformants bearing episomes carrying *A. thaliana* telomere sequences were identified and analyzed.

Southern Hybridizations

Protocols for preparation of total genomic plant DNA were previously described (28). Exonuclease digestion of genomic DNA was performed using 1–5 U/ml of *Bal*31 nuclease at 30°C at a DNA concentration of 7.5 to 10 μg/ml in 12 mM $CaCl_2$, 24 mM $MgCl_2$, 0.2 M NaCl, 20 mM Tris-HCl (pH 8.0), 1 mM EDTA, 250 μg/ml BSA. Radiolabeled probes were prepared by the random priming method (28). Southern blots were prepared using nylon membranes (GeneScreen, NEN) and the hybridization protocols described by Church and Gilbert (29). The filters were washed in 0.2× SSC, 0.1% SDS at 65°C unless otherwise indicated in the figure legends.

DNA Sequencing

The DNA sequences presented in Figures 8 and 9 were determined by standard dideoxy techniques (28) using double-stranded templates and Sequenase™ reagents purchased from United States Biochemical Corporation. Sequencing templates were prepared using an exoIII/S1 nested deletion protocol (28). Some regions were sequenced using oligonucleotide primers purchased from Operon Technologies, Inc. (Alameda, CA). As noted in Figures 8 and 9, only one strand of sequence could be obtained for portions of the telomeric repeat regions due to the poor progression of the polymerase through these simple-sequences.

RESULTS

Isolation of *A. thaliana* telomeres in yeast

To learn more about the molecular organization of *A. thaliana* telomeric regions we isolated telomeric restriction fragments from this plant by selection for DNAs which could function as telomeres, or be recognized as telomeres, in the yeast, *S. cerevisiae*. The cloning strategy is outlined in Figure 1 and detailed in Materials and Methods. Briefly, high molecular weight *A. thaliana* DNA was digested with an endonuclease and circularized under dilute conditions to enrich telomeric restriction fragments, containing one restricted end and one telomeric end, which can not be ligated into a circle. The telomeric restriction fragments were then ligated to a yeast artificial chromosome vector which carries only one telomere. After the chimeric molecules were introduced into yeast cells, transformants were selected and screened for cross-hybridization with an *A. thaliana* telomeric probe ($[TTTAGGG]_n$). Eight transformants carrying plasmids which cross-hybridized to the plant telomeric probe were identified from a total of 1272 transformants. Preliminary restriction analysis indicated that the eight clones represent six distinct genomic sites.

Genomic DNAs from the yeast transformants were examined to determine if they carried linear plasmids with a new telomere derived from an *A. thaliana* telomeric-sequence. Southern blotting experiments showed that the *A. thaliana* telomeric probe hybridized to diffuse bands in all the transformants but did not hybridize to any sequences in the untransformed yeast host (data not shown; See Figure 2). The diffuse bands suggest that the *A. thaliana* telomeric repeats are found at the ends of a linear episome since such hybridization signals are hallmarks of telomeres in yeast, and in most organisms, which possess varying number of telomeric repeats at each chromosomal end.

One of the transformant lines, designated YT1 (BF305-15d[YpAtT1]), was analyzed in more detail to determine if the *A. thaliana* telomeric repeats lie close to a telomere on the yeast episome. YT1 genomic DNA was digested progressively with nuclease *Bal*31 to remove sequences from the ends of linear DNAs, followed by restriction endonuclease digestion and Southern analysis. The behavior of telomeres in such an 'exonuclease-sensitivity' assay is shown in the left panel of Figure 2: a yeast telomeric probe ($[dCA/dTG]_n$ recognizing $[TG_{13}]_n$) hybridized to several diffuse *Xho*I bands (note prominent band of approximately 1.3 kb) which shifted to a faster mobility reflecting their preferential degradation compared to the cross-hybridizing nontelomeric fragments at the top of the gel. Similarly, a Southern blot of *Bal*31-treated YT1 DNA hybridized with the *A. thaliana* telomeric probe (See Figure 2, right panel) showed a diffuse band which shifted mobility in parallel with the yeast telomeric fragments demonstrating that the plant telomeric sequence lies close to one telomere of the linear episome carried in YT1.

The genomic sequences isolated on YpAtT1 and YpAtT7 were derived from chromosomal termini in *A. thaliana*

We focused our characterization on two linear yeast episomes, YpAtT1 and YpAtT7 (carried in yeast transformants, YT1 and YT7, respectively), containing *A. thaliana* telomere cross-hybridizing sequences. To facilitate these studies, we constructed circular plasmid derivatives of YpAtT1 and YpAtT7 which could be propagated in *E. coli* as described in Materials and Methods. The structure of the plant genomic sequences carried on the

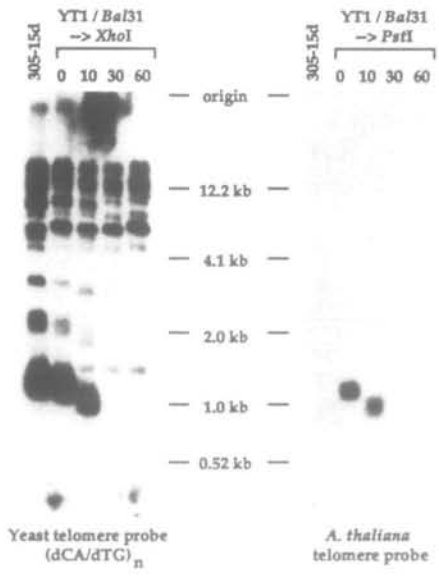

Figure 2. One arm of YpAtT1 carries an *A. thaliana* telomere sequence. Yeast genomic DNA from YT1 (BF305-15d[YpAtT1]) was pretreated with nuclease *Bal*31 for 0, 10, 30 or 60 minutes as indicated to remove increasing amounts of sequence from DNA termini. These DNAs, along with control DNA from the untransformed host (305–15d), were then digested with either *Xho*I (left panel) or *Pst*I (right panel), size-fractionated on an agarose gel and transferred to a nylon membrane. Southern hybridization with a yeast telomere probe (poly dCA/dTG; left panel) demonstrates that the exonucleolytic pretreatment preferentially attacks telomeric fragments; the prominent 1.3 kb *Xba*I band represents the major subclass of yeast telomeres (with Y′ TASs). The Southern blot in the left panel shows that the *A. thaliana* telomeric probe: i) did not hybridize to the untransformed host, and ii) recognized a 1.2 kb exonuclease-sensitive telomeric band carried on YpAtT1. The filters shown in the left panel was washed in 2× SSC, 0.1% SDS at 65°C.

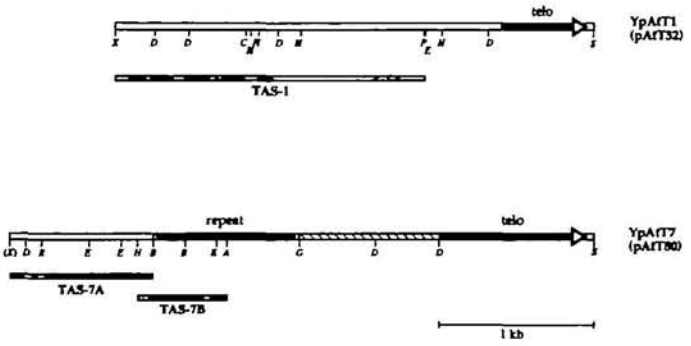

Figure 3. Structure of the YpAtT1 circular derivative, pAtT32, and the YpAtT7 circular derivative, pAtT80. The black boxes on the right represent regions containing plant telomeric repeats ('telo') and the open arrowhead denotes yeast telomeric repeats. The remainder of the inserts correspond to telomere associated sequences (TASs); the open boxes indicate single-copy DNA sequences while the patterned boxes represent repeated DNA sequences. Portions of the TASs used as a hybridization probes in Figures 4, 5, 6 and 7 are indicated by the stippled boxes under the restriction maps. The restriction maps were constructed by: i) partial digestion of endlabeled fragment techniques (28), and ii) reference to DNA sequence data. The restriction sites are abbreviated as follows: *A* = *Apa*I, *B* = *Bam*HI, *C* = *Cla*I, *D* = *Dra*I, *E* = *Eco*RII/*Bst*NI, *G* = *Bgl*II, *H* = *Hpa*II/*Msp*I, *P* = *Pst*I, *R* = *Eco*RI, *S* = *Sac*I, *X* = *Xba*I, and (*X*) = *Xba*I site removed by partial fill-in and ligation (See Figure 1).

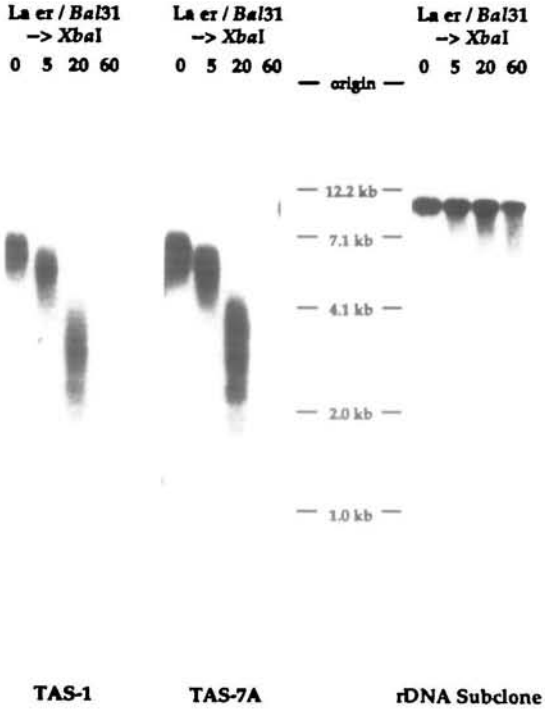

Figure 4. The YpAtT1 and YpAtT7 were derived from chromosomal termini in *A. thaliana*. Landsberg erecta genomic DNA was pretreated with nuclease *Bal*31 for 0, 5, 20 or 60 minutes as indicated at the top of the gel. The DNAs were then digested with *Xba*I, size-fractionated on an agarose gel and transferred to a nylon membrane. The filter was first hybridized with the YpAtT1 TAS1 probe (See Figure 3) demonstrating that the probe was derived from an exonuclease-sensitive telomeric fragment (left panel). Similar results were obtained after rehybridizing the filter with the TAS7A probe from YpAtT7 (center panel). The filter was then rehybridized with a portion of the rDNA as a control for absence of nonspecific degradation of non-telomeric restriction fragments (right panel).

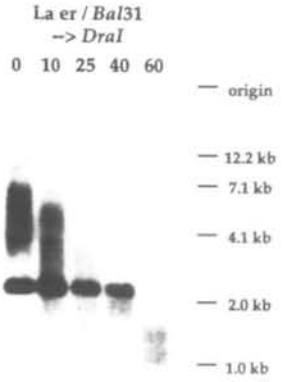

Figure 5. A region of the YpAtT7 TAS is repeated at other telomeres. Landsberg erecta genomic DNA was pretreated with nuclease *Bal*31 for 0, 10, 25, 40 or 60 minutes as indicated at the top of the gel. The DNAs were then digested with *Dra*I, size-fractionated on an agarose gel and transferred to a nylon membrane. The pattern resulting from hybridization with the YpAtT7 TAS7B probe (See Figure 3). The probe recognized exonuclease-sensitive restriction fragments corresponding to the wide smear of 3–7 kb. The discrete 2.1 kb band corresponding to the YpAtT7 insert was also sensitive to exonuclease treatment at later time points.

circular plasmids are shown in Figure 3. The DNA sequences which cross-hybridize to the *A. thaliana* telomeric repeats are denoted by the black boxes at the right end of both inserts. This is the end of the insert which corresponds to the free end of the linear yeast episomes YpAtT1 and YpAtT7 and is the expected position of the terminal telomeric repeats. The remainders of both inserts are composed of both single-copy (open boxes) and repeated DNAs (patterned boxes) which do not cross-hybridize to telomeric repeats.

Since the *A. thaliana* genome contains telomere-similar sequences at non-telomeric locations (30), it was necessary to determine if the *A. thaliana* genomic sequences on YpAtT1 and YpAtT7 were derived from chromosomal termini. DNA probes from the single-copy regions of the clones, indicated as TAS-1 and TAS-7A in Figure 3, were used as hybridization probes on Southern blots of *Bal*31-treated *A. thaliana* Landsberg DNA. The TAS-1 probe hybridized to a diffuse *Xba*I band centered around 6 kb in the lane containing non-*Bal*31-treated DNA (See Figure 4, left panel). This band behaved like a telomeric fragment, shifting to a faster mobility upon increasing *Bal*31 pretreatment. Similarly, the unique TAS-7A probe recognized a co-migrating 6 kb *Xba*I band which was also diffuse and sensitive to exonuclease (Figure 4, center panel). As a negative control, the filter shown in the left and center panel was rehybridized with a nontelomeric probe, a portion of rDNA. As seen in the right panel of Figure 4, the rDNA fragments did not change mobility indicating that random degradation of genomic sequences can not account for the behavior of the TAS cross-hybridizing bands. Moreover, similar blots were hybridized with the *A. thaliana* telomeric repeat probe and several diffuse *Bal*31-sensitive bands were seen, including one which co-migrated with the unique bands seen in Figure 4 (data not shown).

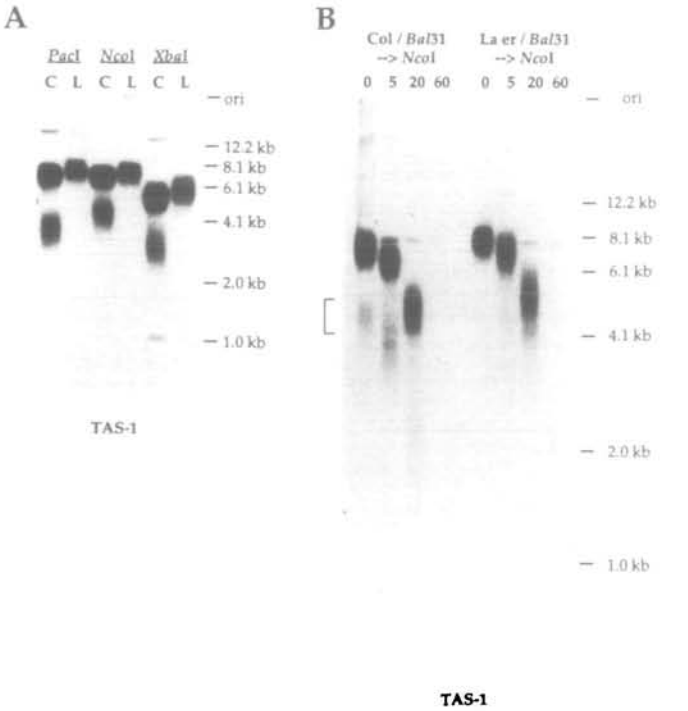

Figure 6. The Landsberg YpAtT1 TAS cross-hybridizes to a second telomere in the *A. thaliana* ecotype Columbia. (A) Genomic DNA from ecotype Columbia (C) and Landsberg erecta (L) were digested with cytosine methylation-insensitive endonucleases (*Pac*I, *Nco*I, *Xba*I) as indicated, size-fractionated on an agarose gel and transferred to a nylon membrane. The filter was hybridized with the Landsberg YpAtT1 TAS1 probe and washed in 0.2× SSC, 0.1% SDS at 60°C. (B) Genomic DNA from ecotype Columbia (Col) and Landsberg erecta (La er) were pretreated with nuclease *Bal*31 for 0, 5, 20 or 60 minutes as indicated at the top of the gel. The DNAs were then digested with *Nco*I, size-fractionated on an agarose gel and transferred to a nylon membrane. The pattern resulted from hybridization with the Landsberg YpAtT1 TAS1 probe (See Figure 3). The bracket indicates the position of the cross-hybridizing telomere in the Columbia ecotype.

These data indicate that the genomic sequences isolated on YpAtT1 and YpAtT7 were derived from chromosomal termini in *A. thaliana*.

The difference between the size of the YpAtT1 and YpAtT7 inserts and the genomic *Xba*I fragments they represent is most likely due to removal of tracts of plant telomere repeats in yeast before addition of the $[TG_{1-3}]_n$ yeast telomere cap. Such processing is well documented in other instances where heterologous telomeres were introduced into yeast (See Discussion). Southern blot experiments demonstrated that the structure of the cloned YpAtT1 and YpAtT7 TASs match that found in the genomic DNA indicating that no deletions or rearrangements of the TASs took place (data not shown).

A region of the YpAtT7 TAS is repeated at other telomeres

The YpAtT7 TAS contains both single-copy sequences, represented by probe TAS7A, and repeated sequences. Lying between the terminal telomeric repeats and the single-copy region are sequences found in multiple copy in the ecotype Landsberg, from which the original telomere library was constructed. Probes from the region abutting the telomeric repeat domain (Figure 3, striped box) hybridized to bands expected from the restriction map of the insert in addition to discrete bands (data not shown). We estimate that the copy number of this region is approximately two to three per genome. Moving away from the telomere we found a region, denoted by the stippled box, containing sequences which also hybridized to additional bands on Southern blots; in this case, the additional hybridization signal corresponded to diffuse bands diagnostic of telomeric restriction fragments. As shown in Figure 5 (lane 0), hybridization with probe TAS-7A detected a broad smear as well as a band corresponding to the YpAtT7 insert on Southern filters containing *Dra*I digested Landsberg genomic DNA. Both the discrete 2.1 kb *Dra*I band and the smear of hybridization were sensitive to exonuclease pretreatment with the expected delayed degradation of the discrete band representing sequences lying several kilobases from a chromosomal terminus. Rehybridization of the filter with the rDNA probe rules out non-specific degradation of genomic sequences (data not shown, provided to referees). These results indicate that the YpAtT7 TAS contains sequences that are found close to at least one other chromosomal terminus. We have been able to resolve the single *Dra*I smear into at least five diffuse bands using the endonuclease *Bst*EII, indicating that the YpAtT7 TAS repeat is found close to several chromosome ends (data not shown).

The Landsberg YpAtT1 TAS cross-hybridizes to a second telomere in ecotype Columbia

While the TAS isolated in YpAtT1 is unique in the *A. thaliana* ecotype Landsberg, Southern blot experiments indicated that the Columbia ecotype contains a second sequence which cross-hybridizes with the TAS-1 probe (See Figure 6A). This second locus appeared as a diffuse band on Southern blots suggesting that the Landsberg YpAtT1 TAS may recognize sequences adjacent to a different telomere in the Columbia ecotype. Southern analysis of *Bal*31-treated Columbia DNA was undertaken to investigate this possibility. Figure 6B demonstrates that the second, weak, diffuse *Nco*I Columbia band (4.5 to 5 kb, denoted by the bracket) is telomeric since it was preferentially degraded in parallel with the darker telomere bands in Columbia and Landsberg. Again, as a negative control, the blot was rehybridized with a rDNA probe to verify that non-telomeric sequences were not degraded (data not shown, provided to referees).

Ecotype differences in telomere length

During the course of these studies, we noticed a consistent difference in the length of telomeric restriction fragments between the two common *A. thaliana* ecotypes, Columbia and Landsberg. This is evident in Figure 6A; the Landsberg telomere detected by probe TAS-1 is larger than the corresponding Columbia telomere. We found no evidence for structural polymorphisms in the TAS corresponding to the YpAtT1 telomere to account for these differences, which we assign to variations in the length of the telomeric repeat tract. The Landsberg telomere is approximately 600 bp longer than the corresponding Columbia telomere (using the center of the dispersed band as a reference point); the difference represents roughly 80−90 repeats of a seven bp telomeric motif.

A. thaliana telomeric regions are methylated

Several of the restriction endonucleases which cut the cloned TASs did not generate the expected fragments when genomic DNA was the substrate. The most likely explanation for this difference is that cytosine methylation masks restriction sites in

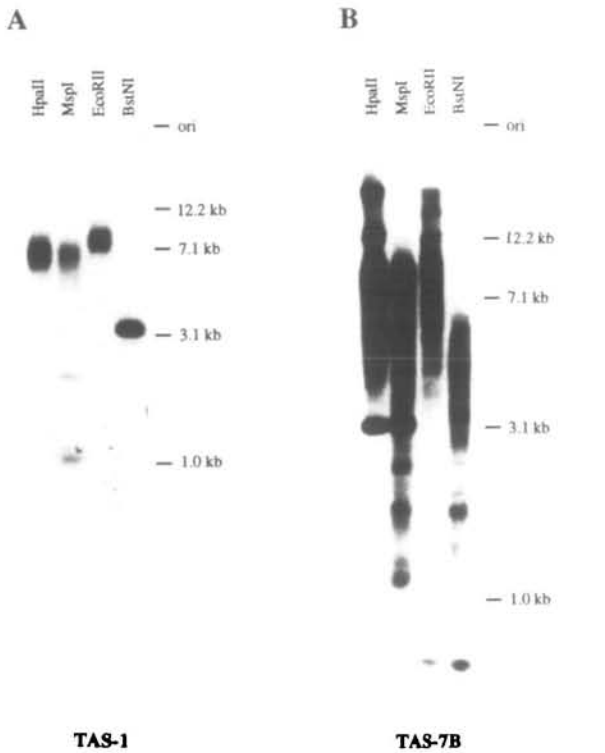

Figure 7. The YpAtT1 and YpAtT7 TASs are methylated in the plant chromosome. Landsberg genomic DNA was digested with the isoschizomer pairs *Hpa*II & *Msp*I, or *Eco*RII & *Bst*NI as indicated at the top of the gel. The DNAs were then size-fractionated on an agarose gel, transferred to a nylon membrane and hybridized with the (A) YpAtT1 TAS1 or (B) YpAtT7 TAS7B probe (See Figure 3). The filter shown in B was washed in 0.2× SSC, 0.1% SDS at 60°C.

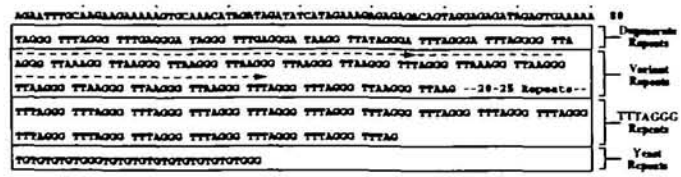

Figure 8. Nucleotide sequence of the distal TAS and telomeric repeat block of YpAtT1. The plant telomeric G-rich repeat block is shown in the boxed regions. The domains described in the text are indicated at the right. The dashed arrows denote a 49 bp repeat within the variant repeat region. The $(TG_{1-3})_n$ telomeric repeats of yeast, added on to the processed plant telomere, are seen at the most distal end of the insert. The TAS extends to nucleotide 80. DNA sequence was obtained from both strands for nucleotides 1 to 31

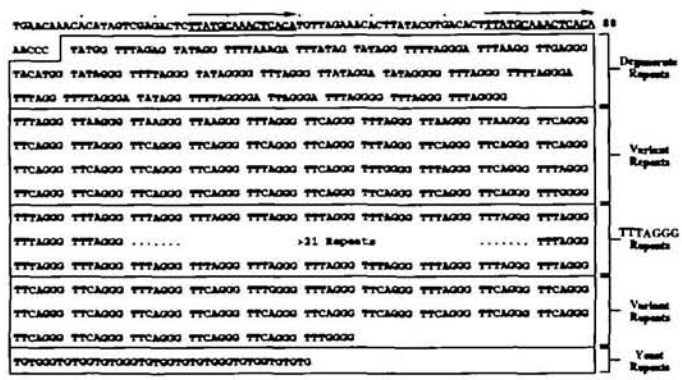

Figure 9. Nucleotide sequence of the distal TAS and telomeric repeat block of YpAtT7. The plant telomeric G-rich repeat block and the $(TG_{1-3})_n$ yeast telomere cap are shown in the boxed regions. The domains described in the text are indicated at the right. The 15 bp direct repeat at the border of the TAS and the telomeric repeats are marked by underlining and the arrows. DNA sequence was obtained from both strands for the entire TAS, as well as the majority of the telomeric repeat region.

the genomic DNA. Methylated cytosines in plant nuclear genomes occur at the sequence CpNpG as well as CpG (31). To search for the presence of methylated cytosines at these sequences we monitored the ability of endonuclease isoschizomer pairs *Eco*RII/*Bst*NI (CCA/TGG) and *Hpa*II/*Msp*I (CCGG) to cleave the genomic TASs. Neither *Eco*RII and *Hpa*II will cut if the second C in their respective recognition sequence is methylated (32).

Consequently, differential cleavage by the two isoschizomers pairs reflects methylation at both CpG (cleavage with *Msp*I but not *Hpa*II) and CpNpG (cleavage with *Bst*NI but not *Eco*RII). In addition, CpNpG methylation can also block *Msp*I digestion as this enzyme will not cut 5mCCGG (32).

The genomic Southern blot shown in Figure 7A suggests that the YpAtT1 TAS is highly methylated at both CpG and CpNpG. As seen in Figure 3, The YpAtT1 TAS has four *Hpa*II/*Msp*I sites and one *Eco*RII/*Bst*NI site. The latter site in the genomic DNA is methylated since no *Eco*RII cleavage was detectable but *Bst*NI cut completely. Similarly, *Hpa*II did not cleave within the genomic TAS indicating that the CCGG sequences are methylated at the second cytosine residue. Moreover, the extremely inefficient cleavage by *Msp*I suggests that the first cytosine in most CCGG sequences is also methylated.

Figure 7B shows a similar Southern analysis of the methylation status of genomic sequences which cross-hybridize to the repeat found in YpAtT7 TAS (probe TAS7B). Again, there was a marked difference between the hybridization signals in the lanes containing DNA digested with isoschizomer pairs indicating the genomic sequences recognized by the probe are heavily methylated.

DNA sequence of the telomeric repeat block and distal TAS region in YpAtT1 and YpAtT7

The DNA sequence of the telomeric repeat block and adjacent TAS from both YpAtT1 and YpAtT7 were determined to learn more about the structure of the telomeric repeat arrays and telomere flanking regions. We note several features in the sequence of the terminal portion of the YpAtT1 insert (distal to the *Pst*I site) displayed in Figure 8. At the most distal end of the block is the $[TG_{1-3}]_n$ yeast telomeric cap added to the ends of heterologous telomeres during propagation in yeast (5, 6, 33, 34). Adjacent to the yeast telomere repeats is a region containing tandem arrays of the previously established TTTAGGG *A. thaliana* telomere repeat sequence. Moving toward the interior of the chromosome, a region containing variant repeats is encountered; the most prominent variant being TTAAGGG. In contrast to the distal $[TTTAGGG]_n$ domain, the structure of the variant repeat region is less regular, with interspersion of TTTAGGG, TTAAGGG and TTAAAGG repeats, although all repeats are of uniform size. The interspersion does not appear

to be random, however, as a higher order organization can be discerned in the variant repeat region (see dashed arrows in Figure 8). The telomeric repeat block decays further towards the centromere-proximal edge of the repeat block which is comprised of several degenerate telomeric repeats of varying size and sequence. Despite this degeneration, a relatively clear boundary can be drawn between the telomeric repeats and the TAS (at nucleotide 80 in Figure 8). The TAS region which abuts the telomeric repeat block is notably AG-rich (81%, nucleotides 10–80). Few additional distinguishing features are present in the remainder of the 0.5 kb of YpAtT1 TAS we have sequenced (EMBL Z12169).

The primary nucleotide sequence of the YpAtT7 telomeric repeat block and adjacent TAS is shown in Figure 9. The telomeric arrays in this clone also display a substructure. Directly internal to the terminal yeast telomere cap is a variant repeat region containing numerous TTCAGGG repeats, followed by region comprised of TTTAGGG repeats. The centromere-proximal boundary of the telomeric repeat block consists of a variant repeat domain which gives way to a stretch of degenerate repeats before the TAS begins. At the junction between the telomeric repeats and the TAS in YpAtT7 we note two 15 bp repeats in direct orientation but few other structural features are evident in the primary sequence of the adjacent 1.7 kb TAS (EMBL Z12170), including the region encompassing the conserved telomere flanking repeat. The YpAtT1 and YpAtT7 TASs do not exhibit any significant similarity to DNA sequences present in the databases, nor do they contain any large open read frames.

DISCUSSION

We have extended our analysis of plant telomeres through study of linear yeast plasmids which carry pieces of *A. thaliana* telomeric and contiguous telomere-associated DNA. Eight clones were isolated which cross-hybridize to *A. thaliana* telomeric repeats after selection for stabilization of yeast centromere plasmids carrying only one telomere. Similar functional complementation schemes, which rely on the structural and functional conservation of telomeres (2, 33, 35), have been used to isolate telomeric DNA from humans (26, 36, 37, 38, 39).

Here, we focused on characterizing two of the eight putative *A. thaliana* telomere clones isolated in yeast and presented evidence that the *A. thaliana* DNAs which serves as telomeres in yeast plasmids YpAtT1 and YpAtT7 were derived from bona fide plant chromosomal ends. Study of these *A. thaliana* telomeric DNAs yields insight into the structure and organization of plant chromosomal termini.

Telomeric Repeat Structure

We previously reported the isolation of a short ($\approx$400 bp) stretch of *A. thaliana* telomeric sequence composed almost entirely of tandem arrays of TTTAGGG repeats (24). In contrast, the telomeric block isolated on YpAtT1 is comprised of subdomains forming a gradient progressing from TTTAGGG repeats in the most distal region to variant repeats and finally degenerate repeats in the centromere-proximal domains. A similar organization is present in the YpAtT7 telomeric repeat block, however, a variant repeat region is also found at a more distal position. We think the substructure seen in the cloned telomeric fragments reflects the organization of the telomeric block *in vivo* but caution must be taken since, as noted earlier, propagation in yeast leads to loss of heterologous telomere sequences and terminal addition of TG_{1-3} repeats (26, 33, 34, 36, 37, 39). This sequence loss could be caused by trimming back the longer plant telomeres in yeast or interstitial deletion of the plant telomere block. Either mechanism would likely preserve the centromere-proximal edge of the telomere block; consequently, we believe the degeneration of the telomere repeats in centromere-proximal regions is real. Support for this conclusion comes from genomic Southern analysis indicating that *MseI* cleavage sites (TTAA), which fall within the common variant repeat, TTAAGGG, are not distributed throughout the telomeric repeat block but are restricted to the centromere-proximal region (data not shown). The presence of variant telomeric repeats clustered around the centromere-proximal border of the telomeric repeat block has also been reported in humans (3, 5, 6, 40). Allshire et al. (3) suggests that the variant repeats may be functional or simply a consequence of the divergence of the centromere-proximal repeats which are buried and not corrected by the normal process of telomere replication which presumably occurs at the end of the repeat block (*e.g.*, by a putative telomerase).

The size of the terminal repeat block of the native telomeres corresponding to the YpAtT1 and YpAtT7 inserts can be estimated from the data presented in Figures 3 and 4. As noted above, there is a size difference between the cloned inserts (YpAtT1=3.0 kb; YpAtT7=3.7 kb) and their cognate 6 kb telomeric *XbaI* fragments which is attributable to loss of plant telomeric repeats in yeast. We calculate that the native telomeres had approximately 3.5 kb of telomeric repeats, taking into account the length of the telomeric sequences which remain on the inserts (YpAtT1 $\approx$0.5 kb; YpAtT7 $\approx$1.0 kb). This figure is larger than our previous estimate of *A. thaliana* telomeric repeat block size, $\leq$2.5 kb (24), and may reflect differences between physiological conditions of the plants used in the different DNA preparations. We have noticed some variation in the size of telomeric restriction fragments even among isogenic sibling progeny (Eric Richards, unpublished), although the consistent ecotype differences in telomere length suggest genetic control of telomere length must be operating as well.

Telomere-Associated Sequences

The main motivation for undertaking the study of telomere-associated regions in *A. thaliana* was to learn how a plant genome is organized at the end of the chromosomal molecule. *A. thaliana* is notable among higher plants for its small genome size (41), currently estimated to be approximately 100 Mb. Like most eukaryotes with small genomes, *A. thaliana* has a genome composed mainly of long stretches single-copy sequences and sequestered blocks of the repeated DNAs (*i.e.*, long-period interspersion) (42). In this light, it is intriguing that the YpAtT1 TAS is single-copy up to the junction with the terminal telomeric simple-sequence block, in contrast to most characterized TASs which are moderately repeated and located in the subterminal regions of different chromosomes. The organization of the YpAtT7 TAS is more characteristic of previously described telomere flanking regions from other organisms, but single-copy sequences are still found less than 2 kb from the edge of the terminal YpAtT7 telomeric repeat block. Further characterization of the isolated telomere clones should indicate if the corresponding chromosomal arms lack a non-genic buffer region between the genic component and the telomere. We would expect, however, that some *A. thaliana* telomere flanking regions will be composed of repetitive DNA based on the presence of C-band positive

heterochromatin at several *A. thaliana* chromosomal termini (43). Perhaps single-copy TASs, such as those described here, will lie between the telomere and more centromere-proximal blocks of repetitive DNA. Evidence for spacers separating the terminal simple-sequence blocks and subtelomeric satellite repeat arrays has been recently presented for tomato chromosomes (44).

Although the YpAtT1 TAS is single-copy in the Landsberg ecotype, it shares a limited similarity with a second TAS in the Columbia ecotype. Similar variable distributions of TASs in yeast and humans have been taken as evidence of genomic rearrangements or exchange of sequence information between different chromosomal termini (5, 6). The weak cross-hybridization to the second Columbia telomere may be caused by reduced sequence similarity or a presence of a short region of high similarity.

The *A. thaliana* YpAtT1 and YpAtT7 TASs exhibit a common feature of telomere-associated regions in humans: a high degree of cytosine methylation (6, 7). The YpAtT1 TAS may be an exception to the normal distribution of methylated cytosine residues in *A. thaliana*, which is heavily skewed towards the repeated DNA component (42). Methylation of the human TASs appears to be under developmental control, as *de novo* methylation of TASs leads to higher levels of methylation in somatic versus germ line cells (6, 7). It will be of interest to determine if comparable developmental alterations in TAS methylation levels occur in *A. thaliana*.

Further investigation of the *A. thaliana* telomere clones isolated here should provide a more complete molecular description of the chromosomal termini of a higher plant. It is also anticipated that the TASs will prove useful as landmarks for the end of the physical and genetic map of *A. thaliana*.

ACKNOWLEDGEMENTS

We thank Bruce Futcher for providing the yeast host, BF305-15d, Joe Colasanti for his gift of pBluescript SK− DNA, and Phil Renna for photography. This work was supported in part by a National Science Foundation grant (DMB-8905589), and a Public Health Service grant from the National Institutes of Health (GM43518). S.C. was supported by funds provided by Burroughs-Wellcome.

REFERENCES

1. Blackburn, E.H. (1991) *Nature*, **350**, 569–573.
2. Blackburn, E.H. (1984) *Cell*, **37**, 7–8.
3. Allshire, R.C., Dempster, M. and Hastie, N.D. (1989) *Nucleic Acids Res.*, **17**, 4611–4627.
4. Chan, C.S.M. and Tye, B.-K. (1983) *Cell*, **33**, 563–573.
5. Brown, W.R.A, Mackinnon, P.J., Villasanté, A., Spurr, N., Buckle, V.J. and Dobson, M.J. (1990) *Cell*, **63**, 119–132.
6. Cross, S., Lindsey, J., Fantes, J., McKay, S., McGill, N. and Cooke, H. (1990) *Nucleic Acids Res.*, **18**, 6649–6657.
7. de Lange, T., Shiue, L., Myers, R.M., Cox, D.R., Naylor, S.L., Killery, A.M. and Varmus, H. (1990) *Mol. Cell. Biol.*, **10**, 518–527.
8. Horowitz, H., Thorburn, P. and Haber, J.E. (1984) *Mol. Cell. Biol.*, **4**, 2509–2517.
9. Zakian, V.A. and Blanton, H.M. (1988) *Mol. Cell. Biol.*, **8**, 2257–2260.
10. Louis, E.J. and Haber, J.E. (1990) *Genetics*, **124**, 533–545.
11. Corcoran, L.M., Thompson, J.K., Walliker, D. and Kemp, D.J. (1988) *Cell*, **53**, 807–813.
12. Louis, E.J. and Haber, J.E. (1990) *Genetics*, **124**, 547–559.
13. Wilkie, A.O.M., Higgs, D.R., Rack, K.A., Buckle, V.J., Spurr, N.K., Fischel-Ghodsian, N., Ceccherini, I., Brown, W.R.A and Harris, P.C. (1991) *Cell*, **64**, 595–606.
14. Carlson, M., Celenza, J.L and Eng, F.J. (1985) *Mol Cell. Biol.*, **5**, 2894–2902.
15. Charron, M.J. and Michels, C.A. (1988) *Genetics*, **120**, 83–93.
16. Venter, U. and Hörz, W. (1989) *Nucleic Acids Res.*, **17**, 1353–1369.
17. Kooter, J M., van der Spek, H.J., Wagter, R., d'Oliveira, C.E., van der Hoeven, F, Johnson, P.J. and Borst, P. (1987) *Cell*, **51**, 261–272.
18. Dunn, B., Szauter, P., Pardue, M.L and Szostak, J.W. (1984) *Cell*, **39**, 191–201.
19. Biessmann, H., Mason, J.M., Ferry, K., d'Hulst, M., Valgeirsdottir, K., Traverse, K.L and Pardue, M.L. (1990) *Cell*, **61**, 663–673.
20. Traverse, K.L. and Pardue, M.L. (1988) *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 8116–8120.
21. Barnes, S.R., James, A.M. and Jamieson, G. (1985) *Chromosoma*, **92**, 185–192.
22. Bedbrook, J.R., Jones, J., O'Dell, M., Thompson, R.D. and Flavell R.B. (1980) *Cell*, **19**, 545–560.
23. Saiga, H. and Edström, J.-E. (1985) *EMBO J.*, **4**, 799–804.
24. Richards, E.J. and Ausubel, F.M. (1988) *Cell*, **53**, 127–136.
25. Vollrath, D., Davis, R.W., Connelly, C. and Hieter, P. (1988) *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 6027–6031.
26. Cheng, J.-F., Smith, C L. and Cantor, C.R (1989) *Nucleic Acids Res.*, **17**, 6109–6127.
27. Allshire, R.C. (1990) *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 4043–4047.
28. Ausubel, F.M., Brent, R., Kingston, R E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) Current Protocols In Molecular Biology John Wiley & Sons, New York.
29. Church, G.M. and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. U.S.A.*, **81**, 1991–1995.
30. Richards, E.J, Goodman, H.M and Ausubel, F.M. (1991) *Nucleic Acids Res.*, **19**, 3351–3357.
31. Gruenbaum, Y., Naveh-Many, T., Cedar, H. and Razin, A. (1981) *Nature*, **292**, 860–862.
32. Nelson, M. and McClelland, M. (1991) *Nucleic Acids Res.*, **19(Suppl.)**, 2045–2071.
33. Pluta, A.F., Dani, G.M., Spear, B.B and Zakian, V.A. (1984) *Proc. Natl. Acad. Sci. U.S.A.*, **81**, 1475–1479.
34. Shampay, J., Szostak, J.W and Blackburn, E.H. (1984) *Nature*, **310**, 154–157.
35. Szostak, J.W. and Blackburn, E.H. (1982) *Cell*, **29**, 245–255.
36. Brown, W.R.A. (1989) *Nature*, **338**, 774–776.
37. Cross, S.H., Allshire, R.C, McKay, S.J., McGill, N.I and Cooke, H.J. (1989) *Nature*, **338**, 771–774.
38. Guerrini, A.M., Ascenzioni, F., Pisani, G., Rappazzo, G., Della Valle, G. and Donini, P. (1990) *Chromosoma*, **99**, 138142.
39. Riethman, H.C., Moyzis, R.K., Meyne, J., Burke, D.T and Olson, M.V. (1989) *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 6240–6244.
40. Cheng, J.-F, Smith, C.L. and Cantor, C.R. (1991) *Nucleic Acids Res.*, **19**, 149–154.
41. Leutwiler, L.W., Hough-Evans, B.R. and Meyerowitz, E.M. (1984) *Mol. Gen. Genet.*, **194**, 15–23.
42. Pruitt, R.E. and Meyerowitz, E.M. (1986) *J. Mol. Biol.*, **187**, 169–183.
43. Schweizer, D., Ambros, P., Gründler, P. and Varga, F. (1987) *Arabid. Inf. Serv.*, **25**, 27–34.
44. Ganal, M.W., Lapitan, N.L.V. and Tanksley, S.D. (1991) *The Plant Cell*, **3**, 87–94.