Organization of Chromosome Ends in *Ustilago maydis*. RecQ-like Helicase Motifs at Telomeric Regions

Patricia Sánchez-Alonso[1] and Plinio Guzmán

Departamento de Ingeniería Genética de Plantas, Centro de Investigación y de Estudios Avanzados del IPN, UNIDAD IRAPUATO, Irapuato, Gto., 36500 México

Manuscript received September 5, 1997
Accepted for publication November 12, 1997

ABSTRACT

In this study we have established the structure of chromosome ends in the basidiomycete fungus *Ustilago maydis*. We isolated and characterized several clones containing telomeric regions and found that as in other organisms, they consist of middle repeated DNA sequences. Two principal types of sequence were found: *UTASa* was highly conserved in nucleotide sequence and located almost exclusively at the chromosome ends, and *UTASb* was less conserved in nucleotide sequence than *UTASa* and found not just at the ends but highly interspersed throughout the genome. Sequence analysis revealed that *UTASa* encodes an open reading frame containing helicase motifs with the strongest homology to RecQ helicases; these are DNA helicases whose function involves the maintenance of genome stability in *Saccharomyces cerevisiae* and in humans, and the suppression of illegitimate recombination in *Escherichia coli*. Both *UTASa* and *UTASb* contain a common region of about 300 bp located immediately adjacent to the telomere repeats that are also found interspersed in the genome. The analysis of the chromosome ends of *U. maydis* provides information on the general structure of chromosome ends in eukaryotes, and the putative RecQ helicase at *UTASa* may reveal a novel mechanism for the maintenance of chromosome stability.

THE ends of eukaryotic chromosomes harbor specialized structures that are essential for chromosome maintenance. These regions consist mostly of repeated DNA sequences, with the ends capped by the telomeres. Telomeres protect and track chromosomes for integrity, permit the complete replication of the terminal regions of the chromosomes, and participate in the correct separation of sister chromatids during mitosis (Zakian 1996; Kirk *et al.* 1997). Telomeres most commonly consist of tandemly repeated short DNA sequences (5 to 8 bp in length), containing clusters of G residues oriented 5′ to 3′ toward the chromosome end. These short repeats are highly preserved in evolution since identical sequences are found in the telomeres of divergent organisms (Zakian 1996). Large repeated sequences of the non-long terminal repeat retroposon/type family also have been found as chromosome ends; these are rare, however, and have been found only at the chromosome ends of Drosophila (Levis *et al.* 1993).

Regulation of telomere length during development and the cell cycle is essential; alteration of telomere length affects cell survival (Greider 1996). Knowledge of the various elements necessary to synthesize and maintain telomeres, and to regulate telomere length is now emerging. One such element is telomerase, a telomere-specific ribonucleoprotein polymerase that synthesizes the telomeric repeats. The RNA component in this enzyme contains sequences complementary to the telomeric repeat for which it is the template (Blackburn 1992). Components of the telomerase have been isolated from various organisms; genes encoding protein components, initially isolated from ciliated protozoa (Collins *et al.* 1995), have some homologs in mammals (Harrington *et al.* 1997; Nakayama *et al.* 1997). A catalytic component of the telomerase, encoding reverse transcriptase motifs, has also been isolated from *S. cerevisiae*, euplotes, *Schizosaccharomyces pombe*, and humans (Lingner *et al.* 1997; Counter *et al.* 1997; Nakamura *et al.* 1997; Meyerson *et al.* 1997). Likewise, the RNA components have been cloned from *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, mammals and from several ciliated protozoa (Singer and Gottschling 1994; McEachern and Blackburn 1995; Blasco *et al.* 1995; Greider 1996). The mechanism controlling telomere length also seems to have common features in nonrelated eukaryotes. For instance, proteins that bind to the telomeric repeats in *S. cerevisiae*, *S. pombe*, and humans, negatively regulate telomere elongation (Marcand *et al.* 1997; Cooper *et al.* 1997; van Steensel and de Lange 1997). These proteins have a similar DNA-binding domain and are thought to operate by preventing elongation of the telomere repeat (Shore 1997).

A complex mixture of middle repeated DNA se-

Corresponding author: Plinio Guzmán, Departamento de Ingeniería Genética de Plantas, Centro de Investigación y de Estudios Avanzados del IPN, UNIDAD IRAPUATO, Apartado Postal 629, Irapuato, Gto., 36500 México. E-mail: pguzman@irapuato.ira.cinvestav.mx

[1]*Present address:* Centro de Investigaciones Microbiológicas, Instituto de Ciencias de la Universidad Autónoma de Puebla, Apartado Postal 1622, Puebla, Pue., México.

quences is commonly found adjacent to the telomeric repeats (Zakian 1996). These sequences, which are known as telomere-associated sequences (TAS), are highly polymorphic in both length and distribution in the genome, and are not as conserved as the short telomeric repeats. There is growing information about the organization of TAS. In *S. cerevisiae*, Plasmodium and humans (Zakian 1996; Dore *et al.* 1990; Brown *et al.* 1990), these sequences consist of a series of middle repeated sequences, which are found in some but not all the chromosomes of the same organism, where they often also display a varying distribution between different chromosomes. In *S. cerevisiae*, two types of telomere associated sequences, called X and Y′, are mainly found. Y′ has been well characterized; it is a highly conserved element present as a single copy or as tandemly repeated copies at the telomeres of some of the chromosomes and it occurs in two classes of 5.3 or 6.7 kb (Chan and Tye 1983a; Chan and Tye 1983b; Zakian and Blanton 1988). X is a composite of sequences less conserved with a core of 475 bp that is present in all chromosome ends (Louis *et al.* 1994; Pryde *et al.* 1995; Louis 1995). *Plasmodium falciparum* TAS contain repetitive elements extending over 60 kb from the telomere repeats. The arrangement of the repetitive elements is preserved in several of the chromosome ends, suggesting a conserved structural organization in this region (Corcoran *et al.* 1988). In humans, TAS contain a diverse array of repeated sequences, which show a polymorphic distribution in the genome and distinct domains have been predicted in such regions (Brown *et al.* 1990; Flint *et al.* 1997). Comparison of TAS shows evidence of structural similarities in such regions among distant species; distinct equivalent domains have been inferred from both human and yeast (Flint *et al.* 1997). Little is known about the function of TAS, although it has been suggested that they have an active role in modulating telomere function. Studies in *S. cerevisiae* indicate that under circumstances of loss of the terminal repeat, TAS are rearranged and amplified within the telomeric regions, restoring telomere function (Lundblad and Blackburn 1993). In *P. falciparum* and other intracellular protozoan parasites, gene conversion between the telomere-proximal antigenic genes has been described as a strategy to enhance genetic variation (Borst and Rudenko 1994; Hernández-Rivas *et al.* 1997). In these species TAS also promote chromosome pairing and facilitate meiotic recombination (de Bruin *et al.* 1994).

We have adopted the fungus *U. maydis*, a basidiomycete that causes the smut disease in maize, as a system for genome organization studies. *U. maydis* possesses various of the advantages of the favorite model *S. cerevisiae* for this type of analysis. The size of the genome and the haploid chromosome number is similar to that of *S. cerevisiae* and molecular genetic approaches to address many biological phenomena are routinely pursued in this fungus (Kinscherf and Leong 1988; Wang *et al.* 1988; Kronstad *et al.* 1989; Fotheringham and Holloman 1989; Bölker *et al.* 1995). We have previously characterized segments of chromosome ends in *U. maydis*. The sequence TTAGGG, which is a common telomeric repeat present in many other eukaryotes, ranging from protozoans to humans, was found tandemly repeated at least 37 times at the chromosome termini in *U. maydis*. In addition, a 376-bp segment of **U**stilago TAS (UTAS) was isolated and shown to be immediately adjacent to the telomeric repeat in many or all of the chromosomes (Guzmán and Sánchez 1994). To pursue our work on the structure of *U. maydis* chromosomes, we have further characterized the terminal regions. Here, we report the identification of two principal types of UTAS in *U. maydis*. In one of them, a novel feature that may have a role in chromosome structure and function was found: an element encoding a helicase showing homology to RecQ helicases.

MATERIALS AND METHODS

Strains and media: *Escherichia coli* strain XL1-blue MRF′[Δ(*mcrA*)183 (*mcrCB-hsdSM-mrr*)173 *recA1 endA1 gyrA96 thi-1 supE44 relA1 lac* (F′ *proAB lacI*^{q} ZΔM15 Tn10)] from Stratagene (La Jolla, CA) was used as host for plasmids and lambda phage clones. LB media supplemented with carbenicillin (50 μg/ml) were used to grow *E. coli* strains. *U. maydis* strain FB2*a2b2* (Banuett and Herskowitz 1989) was provided by Flora Banuett, University of California in San Francisco, and I2, a wild isolate, by Octavio Paredes at CINVESTAV, Unidad Irapuato. *U. maydis* strains were grown in YEPS media (1% yeast extract, 1% peptone, and 1% sucrose) at 30° or in minimal medium (Holliday 1974).

Isolation of clones containing telomeric DNA: A chromosome-end enriched library was constructed and screened as previously described (Guzmán and Sánchez 1994). *Eco*RI-generated restriction fragments of *U. maydis* DNA were selected to generate the library. These DNA fragments were ligated at a high concentration to double digested and alkaline phosphatase-treated *Eco*RI-*Hin*cII Bluescript-KS$^-$ vector (Stratagene), using a 10-fold molar excess of vector to genomic DNA. The ligation mixture was then transformed into *E. coli* XL1-blue *MRF′*. The telomeric repeated sequence $TTAGGG_n$, at a concentration of 5×10^5 cpm/ml, was used as a probe to screen the chromosome-end enriched library. To isolate non-telomeric fragments related to the junction fragments between TAS and the telomeric repeat a genomic λEMBL3 library of *U. maydis* (kindly donated by R. Kahmann) were screened with probe UT1-a (see Figure 1).

DNA manipulation and *Bal*31 sensitivity essays: Procedures for the preparation of DNA from *U. maydis* and for the *Bal*31 treatments were previously described (Guzmán and Sánchez 1994). Southern blotting was done with nylon membranes (Hybond N$^+$; Amersham) and hybridizations were carried out in a 0.5 M Na_2HPO_4 (pH 7.2), 7% SDS, 1 mm EDTA, and 1% BSA solution at a probe concentration of 10^6 cpm/ml. Hybridizations were done at 65° followed by washes with 0.2× SSPE, 0.1% SDS at 65°. For removal of probes a solution of 0.1% SDS and 0.1× SSPE was boiled, poured on the membrane, and allowed to cool to room temperature; this treatment was repeated at least twice.

DNA sequence determination and analysis: The DNA sequence was determined using Sequenase 2.0 (United States Biochemical, Cleveland, OH). Single stranded DNA for se-

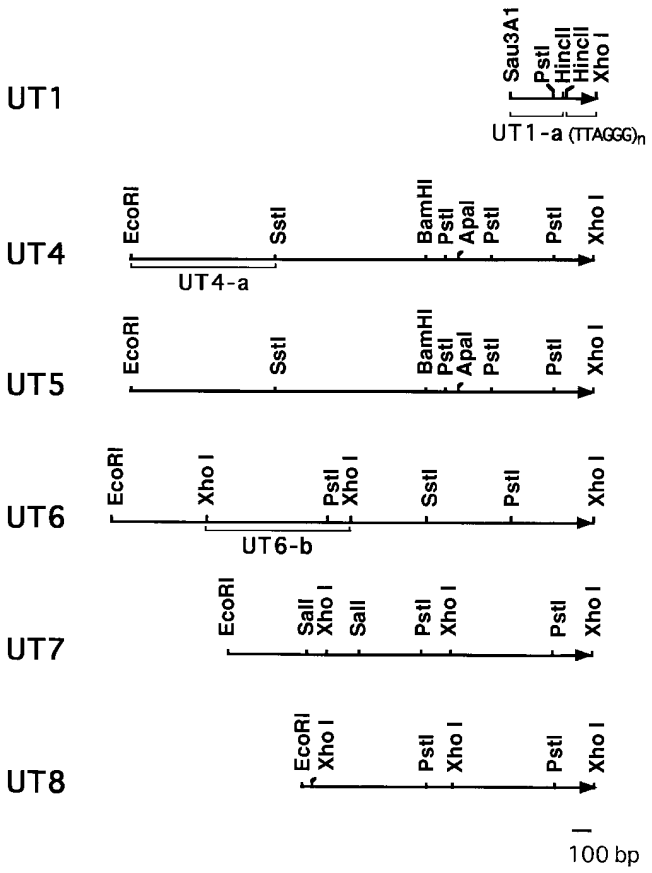


Figure 1.—Restriction maps of five representative clones from *U. maydis* chromosome ends. Clones UT4 to UT8 were isolated from an end-enriched library made with DNA from strain FB2, as described in materials and methods. A series of single and double digests were done to construct the restriction map and the enzymes used for this purpose are indicated; the *Xho*I site at the end is from the polylinker region of the cloning vector. Restriction sites in the map were confirmed after the determination of the entire nucleotide sequence of the fragments. Specific probes for a group of chromosomes are underlined; probe UT4-a is a 785-bp *Eco*RI-*Sst*I fragment and probe UT6-b is a 876-bp *Xho*I-*Xho*I fragment. Other probes used in this study are shown in the upper part; probe TTA GGG_n contains 37 copies of the *U. maydis* telomeric repeat and probe UT1-a, which is a 329-bp *Sau*3A1-*Hin*cII fragment adjacent to the telomeric repeat from clone UT1; UT1-a hybridizes to all the chromosome ends. The arrow head denotes the chromosome end; a 100-bp scale is indicated.

quencing was prepared following the ssDNA template preparation procedure from Promega (Madison, WI). The MacDNASIS ProV 3.2 and DNASTAR packages were used for the analysis of DNA sequences. Homology searches were performed using BLASTX and BLASTP.

Pulsed-field gel electrophoresis (PFGE): Preparation of DNA-agarose plugs from *U. maydis* was carried-out as previously described (Guzmán and Sánchez 1994). The electrophoresis was run at 14° in a CHEF DR III apparatus from Bio-Rad Laboratories (Richmond, CA) using two consecutive blocks. In block 1 pulse times were ramped from 160 to 120 sec for 32 hr at 2.1 V/cm and a 120° angle was used; in block 2 pulse times were ramped from 120 to 80 sec for 24 hr at 5.2 V/cm and a 110° angle was used. Following electrophoresis, gels were stained with ethidium bromide (1 mg/ml) for 30 min, then destained in distilled water. The DNA bands were visualized with a UV transilluminator. *U. maydis* chromosomes were sized by comparison with *S. cerevisiae* chromosomes, purchased from Bio-Rad Laboratories.

RESULTS

Two principal classes of chromosome ends are present in *U. maydis*: In the course of the characterization of telomeric regions from *U. maydis*, we isolated a segment of about 380 bp of a telomere-associated sequence (UTAS), which was found to be immediately adjacent to the telomeric repeats in many or all of the chromosome ends (Guzmán and Sánchez 1994); we speculated that this segment of UTAS would be part of a larger structural component of chromosome ends. Since we were interested in examining in detail the chromosome ends of *U. maydis*, we set out to isolate and characterize a larger segment of UTAS from various chromosomes, which would permit us to define the basic structure of these regions. For this purpose, a library enriched for chromosome ends was constructed from the standard laboratory strain FB2. Twenty-five of about 2400 clones hybridized to telomere repeats. Analysis of 15 of these clones, containing DNA inserts ranging from 1.8 to 2.8 kb, showed that they also hybridized to a region of UTAS located immediately adjacent to the telomeric repeat (probe UT1-a, Figure 1). This result confirmed that the telomeric repeat and the adjacent UTAS were included in the same DNA segment in many of the chromosome ends. The restriction map of five such clones, named UT4 to UT8, and the location of some of the probes used in this study are shown in Figure 1. Using DNA segments from various regions of the cloned UTAS as probes, we found that some of them only hybridized to certain sets of clones (data not shown). Two segments named UT4-a and UT6-b, were selected as probes for further analysis of the chromosome ends; UT4-a hybridized to UT5 but not to UT6, UT7, or UT8, and UT6-b hybridized to UT7 and UT8, but not to UT4 or UT5.

Organization of UT4-a and UT6-b sequences in the *U. maydis* genome: To determine the organization of UTAS we initially carried out a series of Southern blot analyses on three different restriction digests of genomic DNA. Since in previous work genetic variation between *U. maydis* strains was detected using probes derived from chromosome ends, we decided to include in the analysis two strains from a different source (Sánchez-Alonso *et al.* 1996). One strain was FB2, a standard laboratory strain and the other was strain I2, a wild isolate. The two probes from UTAS, UT4-a, and UT6-b, showed different patterns of hybridization (Figure 2). UT4-a detected a prominent signal with DNA from FB2 in each of the three digests and about four signals of minor intensity in two of the digests (Figure 2, left panel). These results suggest that multiple copies

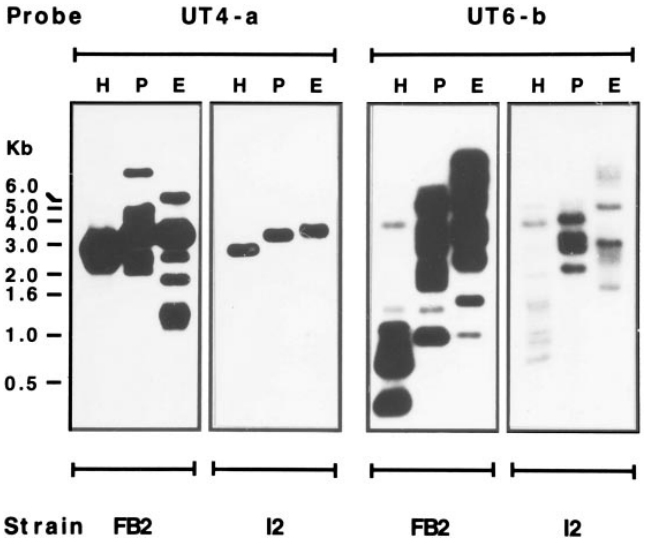


Figure 2.—Analysis of the distribution of UT4-a and UT6-b sequences in the *U. maydis* genome. DNA preparations from strains FB2 and I2 were digested to completion with restriction enzymes: *Hin*cII (H), *Pst*I (P), *Eco*RI (E). DNA was size-fractionated by agarose gel electrophoresis and transferred to a nylon membrane. The blot was hybridized under high stringency conditions to probe UT-4a and then to probe UT-6b after stripping. The removal of the first probe was verified by exposure of the membrane to X-ray film. The positions of the molecular size markers (1 kb ladder from GIBCO BRL, Grand Island, NY) are shown on the left.

of UT4-a, most of them of very similar length, were detected in strain FB2 (see below). In a similar manner, a single signal that seems to coincide in size with the prominent signal observed with FB2, was detected for all digests of DNA from strain I2 (Figure 2, second panel). Analysis of the pattern with UT6-b indicates that this probe detected multiple hybridizing signals in a wide range of sizes and of different intensities in DNA both from strain FB2 and strain I2 (Figure 2, third and fourth panels); in general, the intensity of the signals was stronger for DNA from FB2 than from I2. These Southern blot results suggest that UT4-a and UT6-b are two classes of repeated DNA sequences that show differences in copy number and distribution within the genome as well as between strains.

*Bal*31 sensitivity analysis was then performed to ascertain the location of the probes at the chromosome ends. *U. maydis* DNA was progressively digested for increasing lengths of time with the exonuclease *Bal*31 followed by digestion with *Eco*RI. Southern blot analysis revealed that a shift of the hybridization signal toward a lower molecular mass occurred for two of the signals detected with FB2 when the UT4-a probe was used. The shift occurred in the most prominent signal (3–4-kb region in Figure 3A, left panel) and in one of the minor signals (1.2-kb region in Figure 3A, left panel); the single signal detected for strain I2 was also sensitive to the exonuclease (Figure 3B, left panel). These results suggest

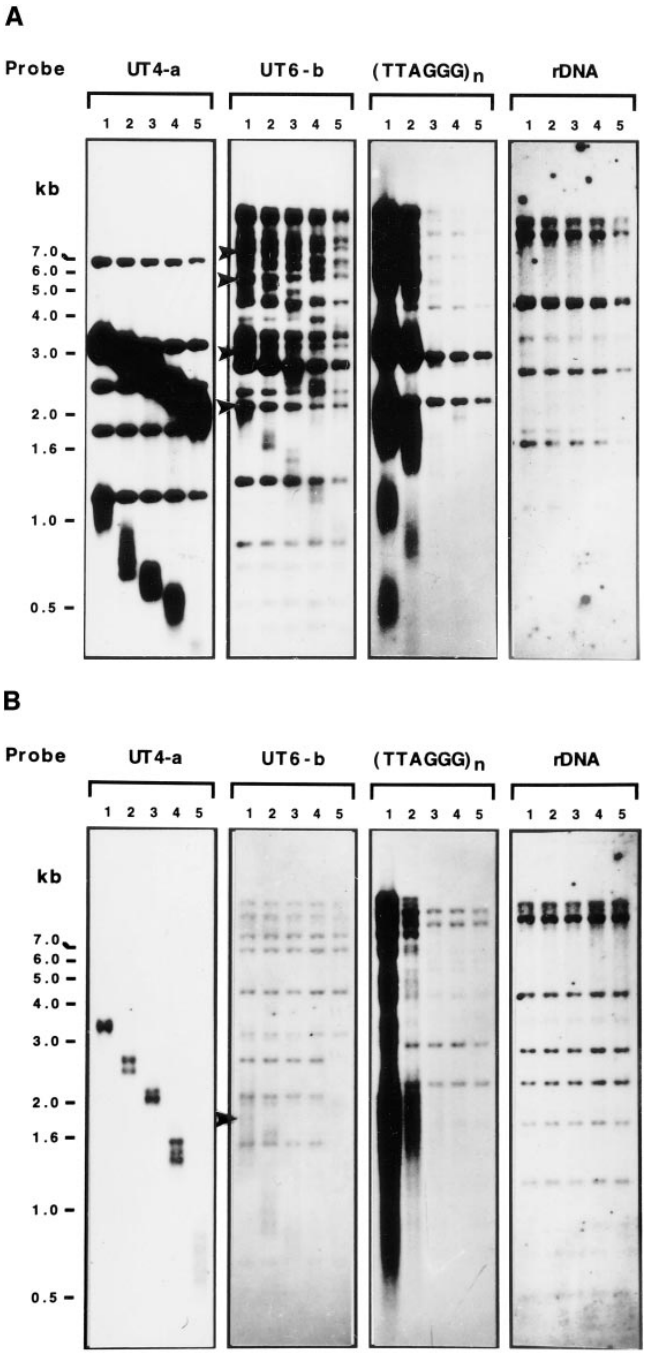


Figure 3.—*Bal*31 sensitivity essay of the telomere-associated sequences UT4-a and UT6-b. DNA preparations from strains FB2 (A) and I2 (B) were digested with the nuclease *Bal*31 for 0 (lane 1), 15 (lane 2), 30 (lane 3), 60 (lane 4), and 120 (lane 5) minutes and subsequently with *Eco*RI. The digested DNA was size-fractionated by agarose gel electrophoresis and transferred to a nylon membrane. Blots were successively hybridized under high stringency conditions with probes UT4-a, UT6-b, $TTAGGG_n$, and rDNA, as indicated; rDNA corresponds to a clone of *U. maydis* ribosomal DNA (P. Sánchez-Alonso and P. Guzmán, unpublished data). Arrowheads point to fragments that show a shift to a lower molecular mass after *Bal*31 treatment. The removal of the probes was verified by exposure of the membrane to X-ray film; the positions of the molecular size markers (1 kb ladder from GIBCO BRL) are shown on the left.

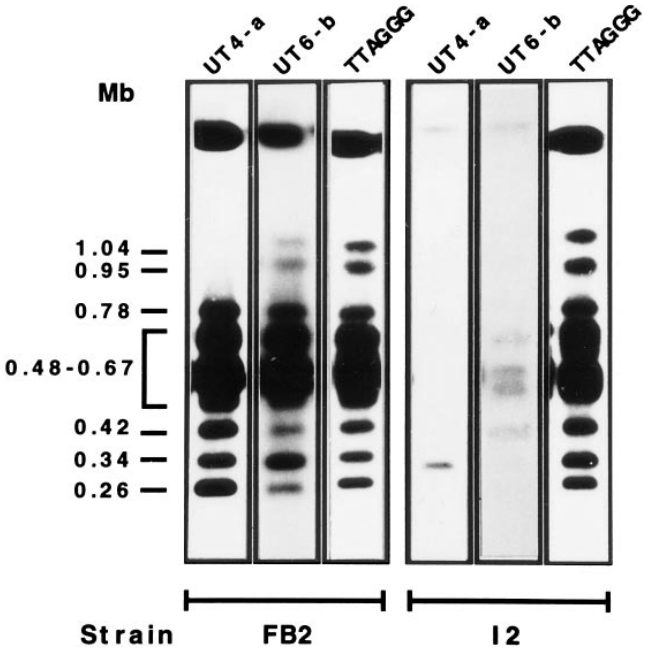


Figure 4.—Location of UT4-a and UT6-b sequences on the *U. maydis* chromosomes. Chromosome preparations of *U. maydis* strains FB2 (first three lanes) and I2 (last three lanes) were subjected to PFGE in a CHEF apparatus; separated chromosomes were transferred to a nylon membrane. Blots were hybridized under high stringency conditions with probes UT4-a, UT6-b, and $TTAGGG_n$, as indicated. The sizes in kilobase pairs of the *U. maydis* chromosomes is shown on the left; they were estimated by comparison to the size of the *S. cerevisiae* chromosomes and to the lambda ladder (GIBCO BRL).

that UT4-a sequences are mainly located at the chromosome ends in both *U. maydis* strains. In addition, strain FB2 contains at least five copies of the sequence that are insensitive to the exonuclease, indicating that they are not located at the chromosome termini. *Bal*31 sensitivity analysis with UT6-b sequences showed a different result. In this case, only a few fragments showed a shift toward the lower molecular mass with FB2 and with I2 (Figure 3, A and B, second panels, pointed by an arrow). This result suggests that most of the fragments detected by UT6-b are interspersed in the genome and not exclusively at the chromosome termini. As controls to assess the proficiency of the *Bal*31 digests, the filters were probed with the telomeric repeat and then with rDNA, a non-telomeric probe. With the first probe, most of the hybridizing signal rapidly disappeared after 10 minutes in both strains (Figure 3, A and B, third panels) and with the non-telomeric probe, the shift toward a lower molecular weight was not observed (Figure 3, A and B, last panels).

The location of UT4-a and UT6-b on the *U. maydis* chromosomes was then determined. Southern blot analysis using these two probes was performed on the electrophoretic karyotype of FB2 and I2 (Figure 4). The two strains displayed differences in the pattern of hybridization. While at least fifteen chromosomes hybridize with the UT4-a probe on DNA from strain FB2 (Figure 4, lane 1), only two showed a signal with DNA from I2 (Figure 4, lane 4; second smallest and largest bands). This observation indicates that most of the signals that UT4-a detects for strain FB2 (Figure 2A, right panel) and for strain I2 (Figure 2B, right panel), consist of various copies of a highly related sequence. Densitometric analysis of the hybridizing chromosomes indicates that the intensity of the signals obtained with UT6-a and the telomeric repeat probes was comparable, suggesting that a similar number of copies of UT4-a are present in each of the hybridizing chromosomes (data not shown). UT6-b detects sequences that were present in most of the chromosomes in the two *U. maydis* strains tested. Densitometric analysis of the hybridizing chromosomes showed differences in the intensity of the hybridizing signals among some of the chromosomes in the same strain when compared to the $TTAGGG_n$ probe (compare the three smallest chromosomes, Figure 4, second and third lanes), suggesting that a different number of copies of UT6-b are present in each of the hybridizing chromosomes (data not shown). This preceding analysis on the organization of UT4-a and UT6-b sequences, and subsequent nucleotide sequence determination, indicate that these two probes are part of two classes of UTAS. Hence forth, we will refer to these two classes of chromosome ends showing homology to UT4-a and UT6-b as *UTASa* and *UTASb*, respectively.

DNA sequence analysis of *UTASa*: DNA sequence analysis of the *UTASa* clones, UT4 and UT5, showed a 90% identity between them in the overall 2.7-kb insert; this result supports the assumption that *UTASa* consists of highly homologous sequences. Some minor differences are observed that correspond to two deletions of fifteen and seven nucleotides in UT5 (positions 1610-1624 and 2694-2700, respectively; see Figure 5) and to four-single nucleotide deletions occurring in the last 400 bp of UT4. Conceptual translation of the sequences revealed open reading frames encoded in the clones that are not altered by the differences observed between the two sequences (see below); these ORFs may be part of longer ORFs whose complete sequence is not included within the cloned fragment.

An ORF containing helicase domains homologous to the RecQ family of DNA helicases is encoded in *UTASa*: Examination of sequence homology of the predicted product of translation found in *UTASa* to sequences in the databases revealed that it encodes a putative helicase. The homology is composed of the seven canonical sequences that are conserved among helicases, including domain I, which corresponds to the A motif of ATPases, domain II, which is probably involved in ATP binding and/or ATP hydrolysis, and domain VI, which may participate in nucleic acid interaction (Gorbalenya *et al.* 1989; Matson *et al.* 1994). This comparison with the databases showed that the best homologies are obtained with RecQ helicases, a particular class of DNA

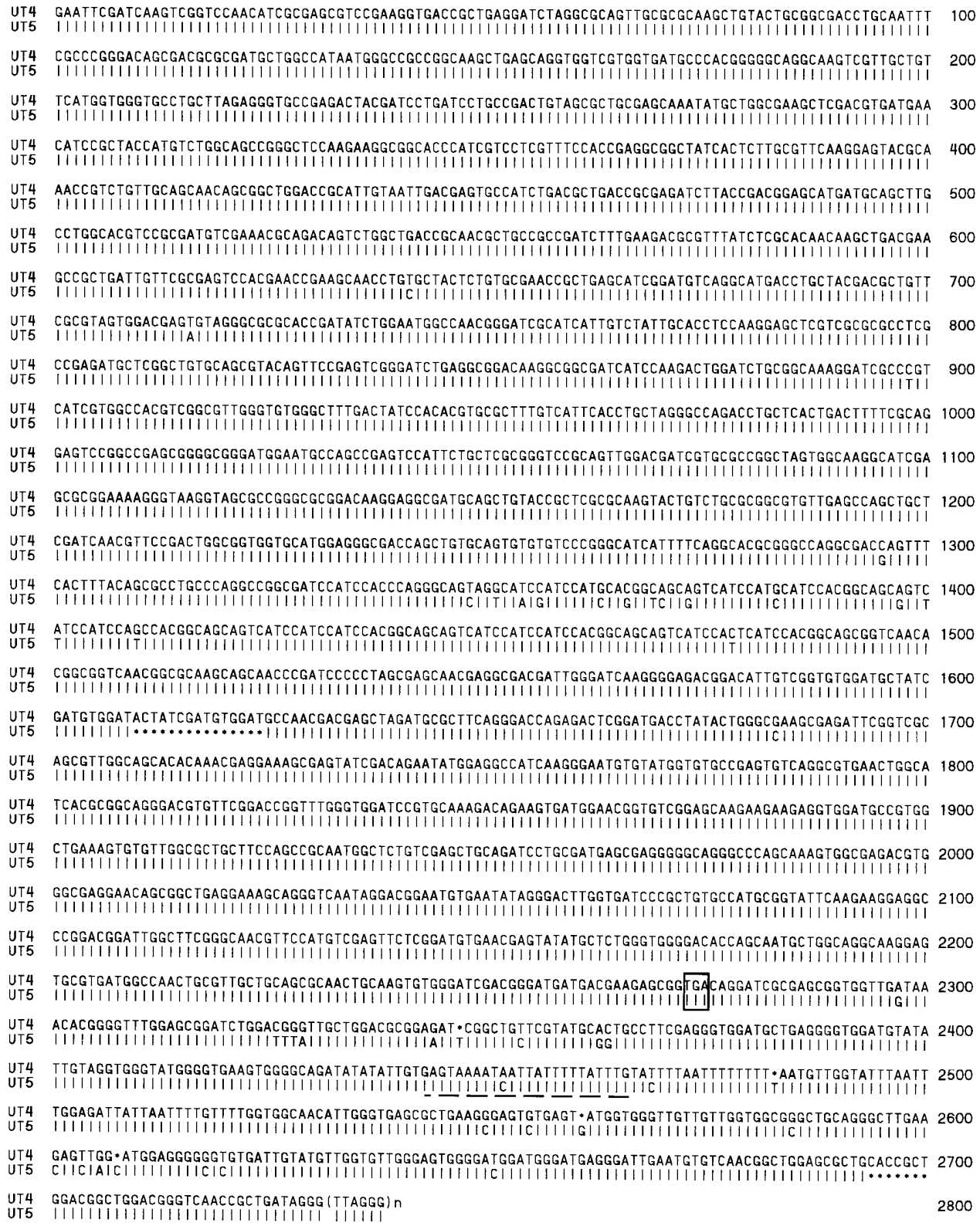


Figure 5.—DNA sequence of *UTASa* chromosome ends. The alignment of the nucleotide sequences of UT4 and UT5 is shown. The sequence is written in a 5′ to 3′ direction towards the chromosome end, which corresponds to the telomeric repeats; identity is indicated by |, a missing nucleotide by (*). Open reading frames of 758 residues in UT4 and of 753 in UT5 initiate at the 5′-end of the sequence and end at the boxed stop codon located at nucleotide 2273. The number on the right indicates the position of the nucleotide residue. Sequences inferred to be at the beginning of a region conserved in almost all chromosome ends are underlined with a broken line. UT4 and UT5 have accession numbers AF030885 and AF030886 in the GenBank database, respectively.

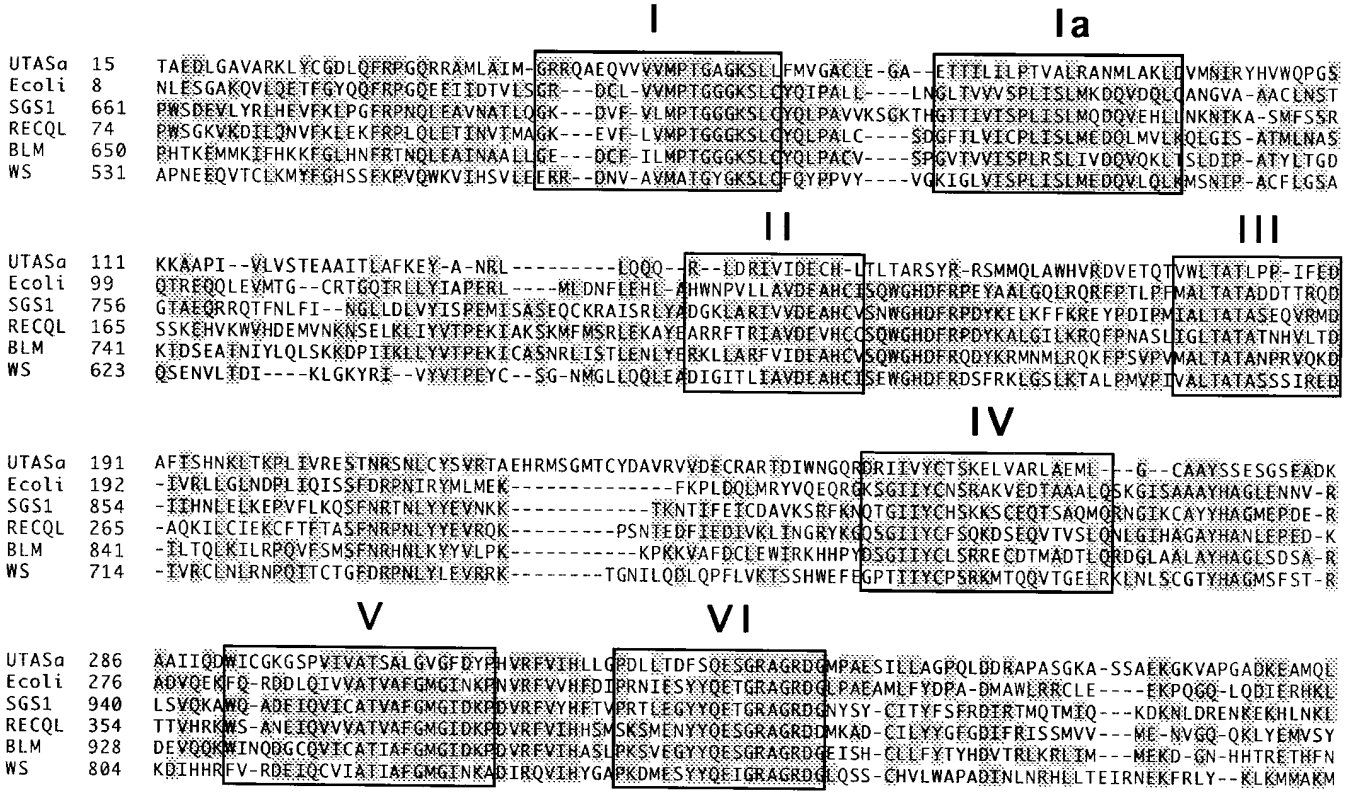
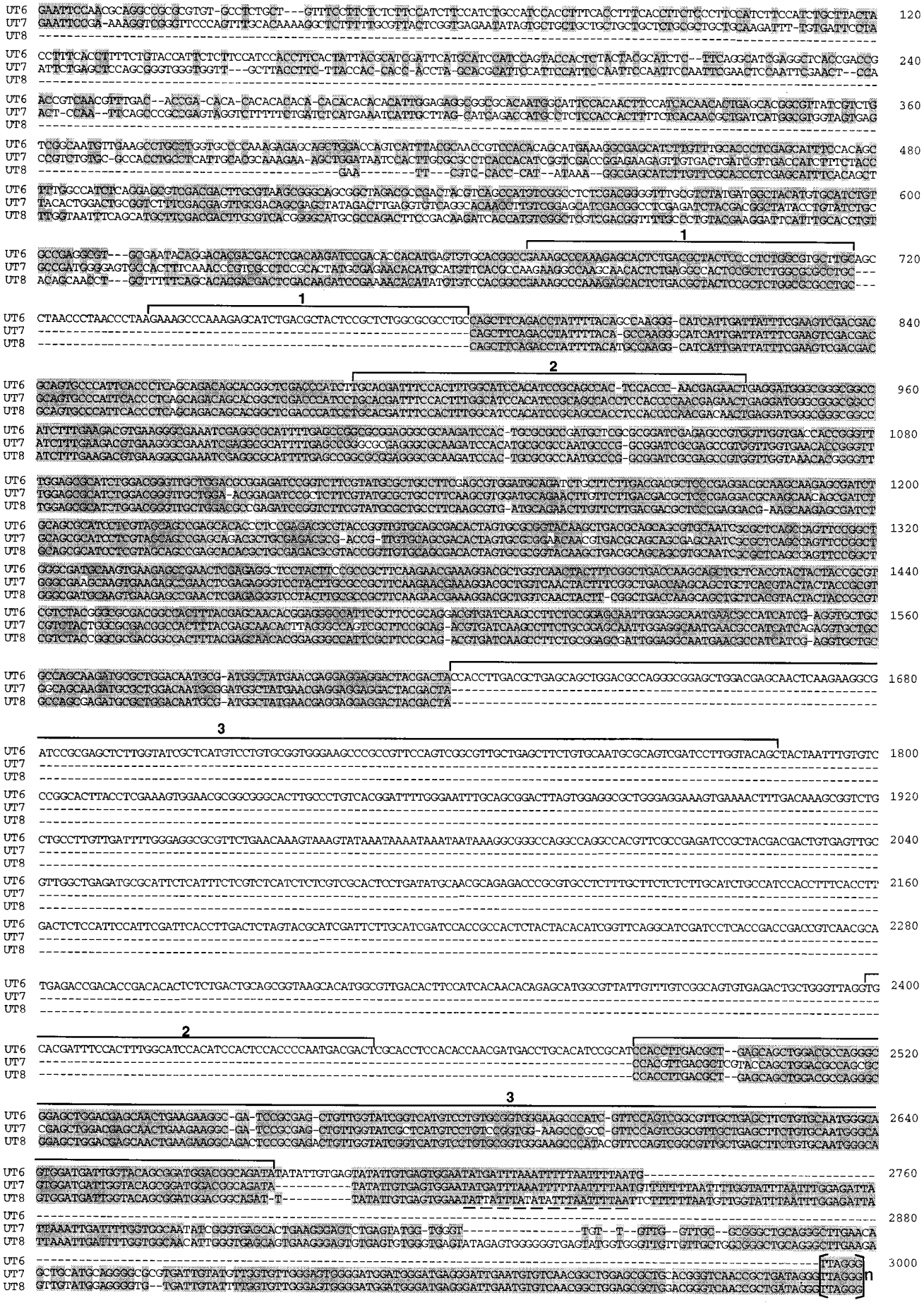


Figure 6.—Helicase domains predicted in *UTASa*. Amino acid alignment of the open reading frame predicted for *UTASa* and the helicase domains of RecQ helicases; the number at the left indicates the position of the amino acid residue for each gene product. RecQ is involved in recombination in *E. coli* (Irino *et al.* 1986); SGS1 is a DNA helicase from *S. cerevisiae* (Gangloff *et al.* 1994); RECQL (Puranam and Blackshear 1994) is human RecQ-like protein and *BLM* (Ellis *et al.* 1995) and *WRN* (Yu *et al.* 1996) are human proteins involved in Bloom's and Werner's syndrome, respectively. Boxed sequences mark the seven helicase domains based on Gorbalenya *et al.* (1989).

helicases that includes gene products from diverse organisms including *E. coli* (Irino *et al.* 1986), *S. cerevisiae* (Gangloff *et al.* 1994; Watt *et al.* 1995), as well as recently described genes involved in Bloom and Werner syndromes in humans (Ellis *et al.* 1995; Yu *et al.* 1996) (Figure 6).

DNA sequence analysis of *UTASb*: The nucleotide sequence of three *UTASb* clones was determined. The insert that hybridized to UT6-b was different in size for each of the clones; about 2.8 kb for UT6, 2.0 kb for UT7 and 1.6 kb for UT8. The DNA sequence analysis revealed that UT7 and UT8 were very similar, having 87% identity in the end-most 1.4 kb of UTAS. The analysis of the nucleotide sequence and of short ORFs detected by the conceptual translation of the sequences (data not shown) do not show homology with previously reported sequences deposited in public databases. When aligned to UT7 and UT8, UT6 showed a high degree of homology in a 837-kb region (position 783-1619, Figure 7). Homology between these clones is interrupted by two DNA insertions in UT6; one is 65 nucleotides long (position 717 to 782, Figure 7) and the other is 865 nucleotides long (position 1620–2485, Figure 7). UT6 also shows three duplicated sequences within the insertion fragments (marked as 1, 2, and 3 in Figure 7), and a 262 nucleotide deletion at the chromosome end when compared to UT8. Interestingly, the sequence CCTAACCCTAACCCTAA (position 718–736, Figure 7), which coincides with complementary telomeric repeats, is found in one of the insertion fragments between repeated sequences; all these features suggest that UT6 has probably undergone rearrangements at UTAS.

UTASa and *UTASb* revealed two types of repeated elements that probably differ in their distribution and conservation in the *U. maydis* genome. We were interested in determining whether these two sequences would be located close to each other in the *U. maydis* genome. We screened a *U. maydis* genomic library in λEMBL3 with UT4-a and UT6-b probes. We found that UT4-a and UT6-b sequences were abundant in the genomic library, being present in about 1% of the clones. Moreover, about 37% of the UT6-b signals hybridized to UT4-a and all of the UT4-a signals hybridized to UT6-b (data not shown). This observation indicates that in most, if not all cases, *UTASa* sequences are associated with *UTASb* sequences in the *U. maydis* genome,


UT6
UT7
UT8
1
2
3
n

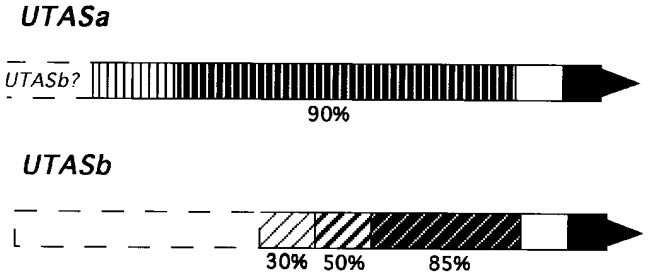


Figure 8.—Schematic illustration of the general organization of the *U. maydis* chromosome ends. Based on DNA sequence analysis of UTAS, two types of chromosome ends, containing *UTASa* and *UTASb*, are predicted. Stippled areas correspond to regions showing different degrees of sequence homology, as indicated. The area corresponding to the telomeric repeats is shown in black and the area common to almost all UTAS is shown in white. *UTASa* may contain *UTASb* sequences toward the centromere (indicated by a noncontinuous box), since all the genomic clones that hybridize to *UTASa* also hybridize to *UTASb*.

and that *UTASb* sequences are also found interspersed in the genome. A general organization of the *U. maydis* chromosome-end regions is depicted in Figure 8.

Identification of UTAS interspersed in the genome: Comparison of the nucleotide sequence of the five UT clones revealed a region highly conserved in all of them. This region is located immediately adjacent to the telomeric repeat and extends for about 300 bp (from the broken line to the end of the sequence on Figures 5 and 7) and in previous work, similar sequences were inferred to occur in non-telomeric regions of the genome (Guzmán and Sánchez 1994). To establish the relationship between the sequences located in telomeric and non-telomeric regions we set out to isolate sequences from non-telomeric regions. We screened a genomic λEMBL3 library of *U. maydis* with a probe containing the conserved segment of UTAS (probe UT1-a, Figure 1). We found that about 1% of the screened clones hybridized to this probe, indicating that as in the case of UT4-a and UT6-b this particular sequence was highly represented in the *U. maydis* genome. Since the λEMBL3 library was generated by *Sau*3A1 partial digestion of DNA, *Sau*3A1 fragments containing telomeric DNA were expected to be under-represented; in fact, no positive clone was found to hybridize to telomeric repeats (data not shown). Restriction digestion and Southern hybridization analysis of ten positive clones revealed a different pattern for each one and a common 200-bp *Pst*I fragment which hybridized to the UTAS probe (data not shown). From four of the clones this 200-bp fragment was cloned and its nucleotide sequence determined; these clones were named UTL1, UTL2, UTL3, and UTL4, for UT-Like sequences. The alignment between these four clones showed a high degree of identity between them over the length of the *Pst*I segment and with a region of UTAS of the UT clones; the homologous region with UTAS was in the *Pst*1 segment adjacent to the telomeric repeat (Figure 9). These observations indicate that a similar type of repeated element is found adjacent to the telomeric repeat in Ustilago chromosomes and interspersed in the genome, suggesting that these sequences have a common origin and that they are important structural components of the *U. maydis* genome. Further examination of the nucleotide sequence of these regions reveals possible sequence duplications that in some of the clones may correspond to tandem arrays (marked as 1, 2, and 3 in Figure 9).

DISCUSSION

The analysis of TAS has been addressed for only a few organisms. These regions of the chromosomes consist commonly of middle repetitive DNA sequences that may vary in their copy number and occurrence in the genome of an individual, and between individuals of the same species. Our analysis of clones containing chromosome ends indicate that this is also the case in *U. maydis*. In this work we described two primary classes of TAS from *U. maydis*, referred to as *UTASa* and *UTASb*. Since genomic clones carrying *UTASa* also hybridize to a probe of *UTASb*, the latter are probably present in all of the chromosome ends. *UTASa*, defines a repeated DNA sequence that is located almost exclusively at the chromosome ends and that is highly conserved in nucleotide sequence. *UTASb* defines a different type of repeated element that is highly abundant and also found interspersed in the genome. Analysis of the end-most regions reveals that break and fusion events have probably occurred in this region. Both classes of UTAS contain a common region of about 300 bp that is located immediately adjacent to the telomere repeats; this sequence is present at almost all of the chromosome-ends and is also found interspersed in the genome. Additional DNA sequence analysis of the chromosome termini and characterization of other chromosome termini will determine the length of the repeated elements described in this work and their structure and organization in the *U. maydis* genome.

Common features between *UTASa* and the *S. cerevisiae*

Figure 7.—DNA sequence analysis of *UTASb* chromosome ends. The alignment of the nucleotide sequences of UT6, UT7, and UT8 is shown. The sequence is written in the 5′ to 3′ direction towards the chromosome end, which corresponds to the telomeric repeats. Position of identity between two or three of the clones is shadowed. The numbers 1, 2, and 3 denote in each case, regions found in clone UT6 that are duplicated. As in Figure 5, sequences inferred to be the beginning of the region conserved in almost all chromosome ends are underlined with a broken line. UT6, UT7, and UT8 have accession numbers AF030887, AF030888, and AF030889 in the GenBank database, respectively.

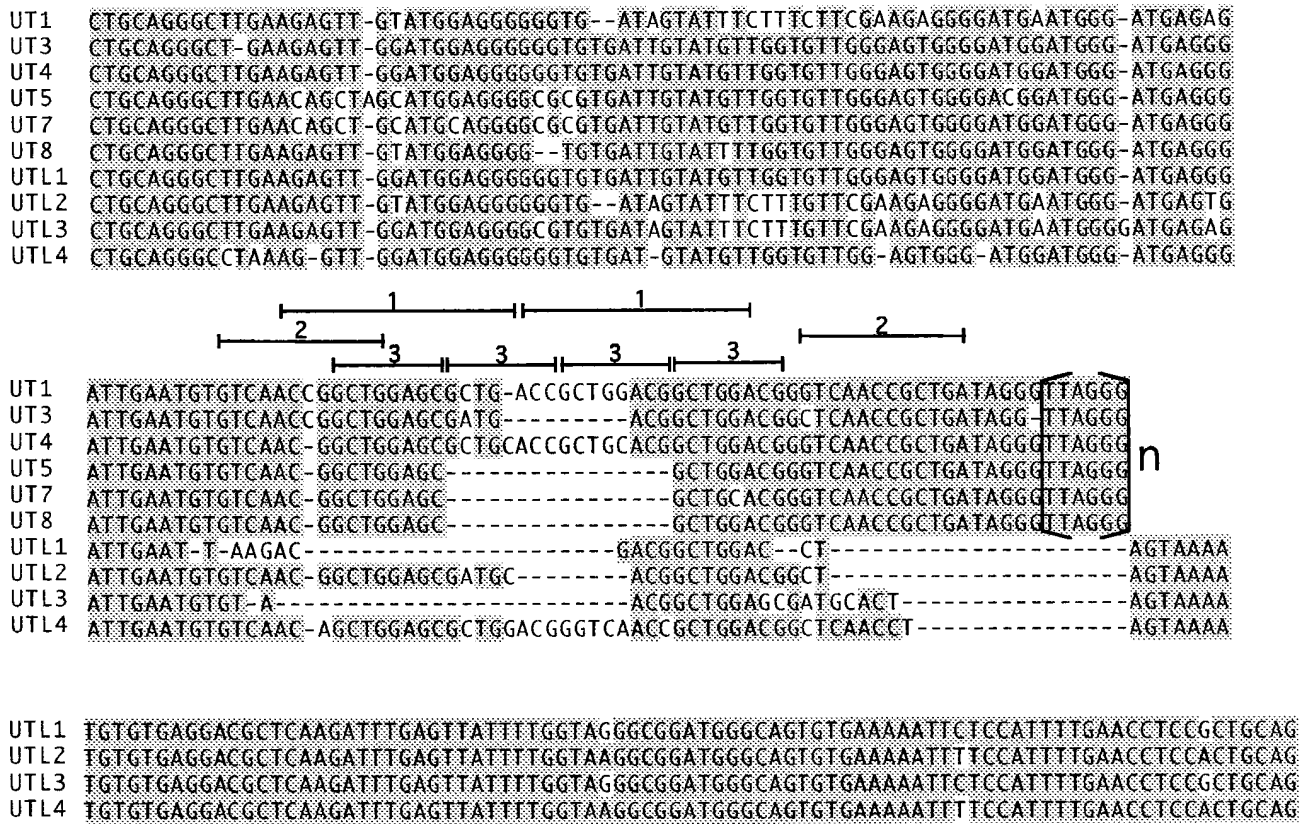


Figure 9.—DNA sequence alignment of UT and UTL clones. The complete sequence of the *Pst*I-*Pst*I fragment from four UTL clones is aligned to the corresponding region in the UT clones; clone UT6 was not included in the alignment because sequence rearrangements have probably occurred in this clone (see text, and Figure 7). The sequence from *Pst*I to the telomeric end from UT-1, UT-3, UT-4, UT-5, UT-7, and UT-8 is shown ending with TTAGGG repeats. Positions of identity in at least 5 of the 10 clones or in at least 2 of the UTL clones are shadowed. The numbers 1, 2, and 3 denote repeated sequences predicted from the alignment. In region 1, 15 out of 18 residues are identical and in region 2, 11 of 12 are identical. Region 3 corresponds to four tandem duplications of an eight bp sequence; the third and fourth repeats are identical and when compared to the first two, 6 out of 8 nucleotides are identical. UTL1, UTL2, UTL3, and UTL4 have accession numbers AF030890, AF030891, AF030892, and AF030893 in the GenBank database, respectively.

Y′ sequences can be predicted. For instance, the *UTASa* element is highly conserved in nucleotide sequence, it is preferentially located at the chromosome ends, it is found in some but not all chromosomes and it varies in copy number and chromosome location among strains. These are all features shared with the *S. cerevisiae Y′* sequences (Zakian and Blanton 1988; Louis and Haber 1992). Another significant observation is that both Y′ and the *UTASa* encode an ORF and for both, the predicted amino acid sequence revealed the seven conserved domains found in helicases (Louis and Haber 1992). No significant homology of the helicase at Y′ to the helicase present at UTAS was detected. However, these helicases may be related, since both show the DExH box in domain II with the corresponding variation QxxGRxxR in domain VI. This reciprocity has been previously suggested in helicases; polypeptides with the DExD box show the sequence HxxGRxxR in domain VI whereas those presenting DExH show the sequence QxxGRxxR (Gorbalenya *et al.* 1989). All these similarities provide evidence that *UTASa* and *Y′* may have equivalent roles in the *U. maydis* and *S. cerevisiae* genomes. The finding of a similar type of element at the chromosome termini of these two fungi is a meaningful feature, suggesting that the location of helicases at the chromosome ends may have structural and functional significance. The finding of a polymorphic distribution of *UTASa* sequences and variation in copy number of this sequence among strains, indicates that the helicase is not essential for the functioning of individual chromosomes. One can speculate that this type of chromosome end may have an advantage for chromosome function and that common mechanisms may participate in structuring this type of telomeric region. It would be interesting to determine whether elements with the properties of *Y′* and *UTASa* are present at the chromosome ends of other organisms.

The helicase found at *UTASa* is most closely related to the RecQ family of DNA helicases. Analysis of the phenotypes associated with recQ helicases suggest that the function of this type of helicase is conserved from bacteria to mammals. Members of this family that have

been isolated are likely to be involved in maintenance of genome stability; these are the *S. cerevisiae SGS1* gene (Watt *et al.* 1996) and the human *BLM* and *WRN* genes, for Bloom's and Werner's syndromes, respectively (Ellis *et al.* 1995; Yu *et al.* 1996). An abnormal increase of mitotic recombination is observed in *S. cerevisiae sgs1* mutants and in Bloom's syndrome cell lines; an abnormal replication and diverse levels of chromosomal translocations and deletions are observed in Werner's syndrome cell lines. The function of the *E. coli recQ* gene correlates with these observations, since it has been shown to be a suppressor of illegitimate recombination (Hanada *et al.* 1997). Premature aging phenotypes are also associated with alterations in some of these helicases; individuals with the Werner's syndrome show several symptoms of premature aging and the *S. cerevisiae sgs1* mutants display alterations that can be correlated with aging disorders (Sinclair *et al.* 1997). It could be speculated that the helicases encoded at *UTASa* may have a role in maintaining genome stability. A single helicase gene expressed at a given time might be sufficient for this function. In such a case, the distinct location of this recQ-like helicase at chromosome ends may be of functional importance. The function of the helicase at UTASa may be related to that of SGS1 and WRN gene products, but having a specific mechanism to regulate its expression. The UTASa helicase may be subject to a telomere-position effect, with its expression being modulated by the silencing mechanism occurring at the telomere region; such a mechanism would not function under normal conditions, but would be turned on under circumstances of defective telomere maintenance. It has been shown in Drosophila, *S. cerevisiae*, *S. pombe*, and trypanosoma genes that are transcriptionally active in an endogenous location of the chromosome become transcriptionally inactive when placed near the telomere (Levis *et al.* 1985; Gottsching *et al.* 1990; Nimmo *et al.* 1995; Horn and Cross 1995).

In conclusion, the analysis of telomere associated sequences in *U. maydis* provides information on the structure and potential functions of these chromosomal regions. A novel feature that may be important for chromosome function is predicted; a putative helicase of the RecQ family. These observations are consistent with the general knowledge that the structure of TAS consists of middle repeated sequences and their involvement in recombination events. Further characterization of the putative RecQ-like helicase present in *UTASa* will reveal more information on the possible role of UTAS in a mechanism to ensure chromosome stability and on its role structuring telomeric regions.

We thank Flora Banuett and Octavio Paredes for Ustilago strains, Regine Kahmann for the Ustilago genomic library in λEMBL3, Beatriz Jiménez and Guillermo Corona for DNA sequencing, Cristina Garcidueñas-Piña for technical assistance, and Gabriela Olmedo, June Simpson, William K. Holloman, and an anonymous reviewer for critical analysis of the work and comments on the manuscript; P.S. acknowledges a scholarship from CONACyT México. This work was supported in part by grants from CONACyT, México to P.G.

LITERATURE CITED

Banuett, F., and I. Herskowitz, 1989 Different *a* alleles of *Ustilago maydis* are necessary for maintenance of filamentous growth but not for meiosis. Proc. Natl. Acad. Sci. USA **86:** 5878–5882.

Blackburn, E. H., 1992 Telomerases. Annu. Rev. Biochem. **61:** 113–129.

Blasco, M. A., W. Funk, B. Villeponteau and C. W. Greider, 1995 Functional characterization and developmental regulation of mouse telomerase RNA. Science **269:** 1267–1270.

Bölker, M., H. U. Böhnert, K. H. Braun, J. Görl and R. Kahmann, 1995 Tagging pathogenicity genes in *Ustilago maydis* by restriction enzyme-mediated integration (REMI). Mol. Gen. Genet. **248:** 547–552.

Borst, P., and G. Rudenko, 1994 Antigenic variation in African *Trypanosomes.* Science **264:** 1872–873.

Brown, W. R. A., P. J. McKinnon, A. Villasanté, N. Spurr, V. J. Buckle *et al.*, 1990 Structure and polymorphism of human telomere-associated DNA. Cell **63:** 119–132.

Chan, C. S. M., and B.-K. Tye, 1983a Organization of DNA sequences and replication origins at yeast telomere. Cell **33:** 563–573.

Chan, C. S. M., and B.-K. Tye, 1983b A family of *Saccharomyces cerevisiae* repetitive autonomously replicating sequences that have very similar environments. J. Mol. Biol. **168:** 505–523.

Collins, K., R. Kobayashi and C. W. Greider, 1995 Purification of *Tetrahymena* telomerase and cloning of genes encoding the two protein components of the enzyme. Cell **81:** 677–686.

Cooper, J. P., E. R. Nimmo, R. C. Alshire and T. R. Cech, 1997 Regulation of telomere length and function by a Myb-domain protein in fission yeast. Nature **385:** 744–747.

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.

Counter, C. M., M. Meyerson., E. N. Eaton and R. A. Weinberg, 1997 The catalytic subunit of yeast telomerase. Proc. Natl. Acad. Sci. USA **94:** 9202–9207.

de Bruin, D., M. Lanzer and J. V. Ravetch, 1994 The polymorphic subtelomeric regions of *Plasmodium falciparum* chromosomes contain arrays of repetitive sequence elements. Proc. Natl. Acad. Sci. USA **91:** 619–623.

Dore, E., T. Pace, M. Ponzi, L. Picci and C. Frontali, 1990 Organization of subtelomeric repeats in *Plasmodium berghei*. Mol. Cell. Biol. **10:** 2423–2427.

Ellis, N. A., J. Groden, T. Ye, J. Straughen, D. J. Lennon *et al.*, 1995 The Bloom's syndrome gene product is homologous to *RecQ* helicases. Cell **83:** 655–666.

Fint, J., G. P. Bates, K. Clark, A. Dorman, D. Willingham *et al.*, 1997 Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. Human Mol. Genet. **8:** 1305–1314.

Fotheringham, S., and W. K. Holloman, 1989 Cloning and disruption of *Ustilago maydis* genes. Mol. Cell. Biol. **9:** 4052–4055.

Gangloff, S., J. P. McDonald, C. Bendixen, L. Arthur and R. Rothstein, 1994 The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase. Mol. Cell. Biol. **14:** 8391–8198.

Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko and V. M. Blinov, 1989 Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res. **17:** 4713–4729.

Gottsching, D. E., O. M. Aparicio, B. L. Billington and V. A. Zakian, 1990 Position effect at *S. cerevisiae* telomeres: reversible repression of PolII transcription. Cell **63:** 751–762.

Greider, C. W., 1996 Telomere length regulation. Annu. Rev. Biochem. **65:** 337–365.

Guzmán, P. A., and J. G. Sánchez, 1994 Characterization of telomeric regions from *Ustilago maydis*. Microbiology **140:** 551–557.

Hanada, K., T. Ukita, Y. Kohno, K. Saito, J. Kato *et al.*, 1997 RecQ

DNA helicase is a suppressor of illegitimate recombination in *E. coli*. Proc. Natl. Acad. Sci. USA **94:** 3860–3865.

Harrington, L., T. McPhail, V. Mar, W. Zhou, R. Oulton *et al.*, 1997 A mammalian telomerase-associated protein. Science **275:** 973–977.

Hernández-Rivas, R., D. Mattei, Y. Sterkers, D. S. Peterson, T. E. Wellems *et al.*, 1997 Expressed *var* genes are found in *Plasmodium falciparum* subtelomeric regions. Mol. Cell. Biol. **17:** 604–611.

Holliday, R., 1974 *Ustilago maydis*, pp. 575–595 in *Handbook of Genetics*, edited by R. C. King. Plenum Press, New York.

Horn, D., and G. A. Cross, 1995 Developmental regulated position effect at a telomeric locus in *Trypanosoma brucei*. Cell **83:** 555–561.

Irino, N., K. Nakayama and H. Nakayama, 1986 The *recQ* gene of *Escherichia coli* K12: Primary structure and evidence for SOS regulation. Mol. Gen. Genet. **205:** 298–304.

Kinscherf, T. G., and S. A. Leong, 1988 Molecular analysis of the karyotype of *Ustilago maydis*. Chromosoma **96:** 427–433.

Kirk, K. E., B. P. Harmon, I. K. Reichardt, J. W. Sedat and E. H. Blackburn, 1997 Block in anaphase chromosome separation caused by a telomerase template mutation. Science **275:** 1478–1481.

Kronstad, J. W., S. F. Covert, D. W. Holden, G. L. McKnight and S. A. Leong, 1989 Isolation of metabolic genes and demonstration of gene disruption in the phytopathogenic fungus *Ustilago maydis*. Gene **79:** 97–106.

Levis, R. W., R. Ganesan, K. Houtchens, L. A. Tolar and F. Sheen, 1993 Transposons in place of telomeric repeats at a *Drosophila* telomere. Cell **75:** 1083–1093.

Levis, R., T. Hazelrigg and G. M. Ruben, 1985 Effect of genome position on the expression of transduced copies of the *white* gene of Drosophila. Science **229:** 558–561.

Lingner, J., T. R. Hughes, A. Shevchenko, M. Mann, V. Lundblad *et al.*, 1997 Reverse transcriptase motifs in the catalytic subunit of telomerase. Science **276:** 561–566.

Louis, E. J., 1995 The chromosome ends of *Saccharomyces cerevisiae*. Yeast **11:** 1553–1573.

Louis, E. J., and J. E. Haber, 1992 The structure and evolution of subtelomeric Y′ repeats in *Saccharomyces cerevisiae*. Genetics **131:** 559–574.

Louis, E. J., E. S. Naumova, A. Lee, G. Naumov and J. E. Haber, 1994 The chromosome end in yeast: its mosaic nature and influence on recombinational dynamics. Genetics **136:** 789–802.

Lundblad, V., and E. H. Blackburn, 1993 An alternative pathway for yeast telomere maintenance rescues *est1*$^-$ senescence. Cell **73:** 347–360.

Marcand, S., E. Gilson and D. Shore, 1997 A protein-counting mechanism for telomere length regulation in yeast. Science **275:** 986–990.

Matson, S. W., D. W. Bean and J. W. George, 1994 DNA helicases: enzymes with essential roles in all aspects of DNA metabolism. BioEssays **16:** 13–21.

McEachern, M. J., and E. H. Blackburn, 1995 Runaway telomere elongation caused by telomerase RNA gene mutations. Nature **376:** 403–409.

Meyerson, M., C. M. Counter, E. N. Eaton, L. W. Ellisen, P. Steiner *et al.*, 1997 hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization. Cell **90:** 785–795.

Nakamura, T. M., G. B. Morin, K. B. Chapman, S. L. Weinrich, W. H. Andrews *et al.*, 1997 Telomerase catalytic subunit homologs from fission yeast and human. Science **277:** 955–959.

Nakayama, J., M. Saito, H. Nakamura, A. Matsuura and F. Ishikawa, 1997 *TLP1*: a gene encoding a protein component of mammalian telomerase is a novel member of WD repeats family. Cell **88:** 875–884.

Nimmo, E. R., G. Cranston and R. C. Allshire, 1994 Telomere-associated chromosome breakage in fission yeast results in variegated expression of adjacent genes. EMBO J. **13:** 3801–3811.

Pryde, F. E., T. C. Huckle and E. J. Louis, 1994 Sequence analysis of the right end of chromosome XV in *Saccharomyces cerevisiae*: an insight into the structural and functional significance of subtelomeric repeat sequences. Yeast **11:** 371–382.

Puranam, K. I., and P. J. Blackshear, 1994 Cloning and characterization of RECQL, a potential human homologue of the *Escherichia coli* DNA helicase RecQ. J. Biol. Chem. **269:** 29838–29845.

Sánchez-Alonso, P., M. E. Valverde, O. Paredes-López and P. Guzmán, 1996 Detection of genetic variation in *Ustilago maydis* strains by probes derived from telomeric sequences. Microbiology **142:** 2931–2936.

Shore, D., 1997 Telomeres: Different means to common ends. Nature **385:** 676–677.

Sinclair, D. A., K. Mills and L. Guarente, 1997 Accelerated aging and nucleolar fragmentation in yeast *sgs*1 mutants. Science **277:** 1313–1316.

Singer, M. S., and D. E. Gottschling, 1994 *TLC1*: template RNA component of *Saccharomyces cerevisiae* telomerase. Science **266:** 404–409.

van Steensel, B., and T. de Lange, 1997 Control of telomere length by the human telomeric protein TRF1. Nature **385:** 740–743.

Wang, J., D. W. Holden and S. A. Leong, 1988 Gene transfer system for the pathogenic fungus *Ustilago maydis*. Proc. Natl. Acad. Sci. USA **85:** 865–869.

Watt, P. M., I. D. Hickson, R. H. Borts and E. J. Louis, 1996 *SGS1*, a homologue of the Bloom's and Werner's syndrome genes, is required for maintenance of genome stability in *Saccharomyces cerevisiae*. Genetics **144:** 935–945.

Watt, P. M., E. J. Louis, R. H. Borts and I. D. Hickson, 1995 *SGS1:* a eukaryotic homolog of *E. coli recQ* that interacts with topoisomerase II in vivo and is required for faithful chromosome segregation. Cell **81:** 253–260.

Yu, C., J. Oshima, Y. Fu, E. M. Wijsman, F. Hisama *et al.*, 1996 Positional cloning of the Werner's syndrome gene. Science **272:** 258–262.

Zakian, V. A., 1996 Structure, function, and replication of *Saccharomyces cerevisiae* telomeres. Annu. Rev. Genet. **30:** 141–172.

Zakian, V. A., and H. M. Blanton, 1988 Distribution of telomere-associated sequences on natural chromosomes of *Saccharomyces cerevisiae*. Mol. Cell. Biol. **8:** 2257–2260.

Communicating editor: S. Jinks-Robertson