

CLONING AND CHARACTERIZATION OF SUBTELOMERIC REGIONS OF

*KLUYVEROMYCES LACTIS*

by

KRISTY ANN PHILLIPS NICKLES

(Under the Direction of Michael J. McEachern)

ABSTRACT

Telomeres are the specialized structures at the ends of eukaryotic chromosomes and are composed of short T/G-rich DNA repeats and the proteins that interact with them. Internal to telomeres are subtelomeric regions that are species-specific and often repetitive. The yeast *Kluyveromyces lactis* has telomeric tracts of 10-20 copies of a 25-bp repeat, but the subtelomeric regions have not previously been characterized in detail. Here we have cloned and characterized subtelomeric regions from ten of the twelve chromosome ends. The amount of sequence examined ranged 0.7-10 kb for each subtelomeric region. We have identified a *K. lactis* subtelomeric element, the R element, which has a purine/pyrimidine strand bias and extends for about 2 kb. Internal to the R element, we found extensive homology among half of the chromosome ends reported here. This homology appears to include three putative gene families.

INDEX WORDS: *Kluyveromyces lactis*, subtelomeric region, telomere, chromosome ends, homology, gene families

CLONING AND CHARACTERIZATION OF SUBTELOMERIC REGIONS OF

*KLUYVEROMYCES LACTIS*

by

KRISTY ANN PHILLIPS NICKLES

B.S., Clemson University, 1999

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment

of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2003

© 2003

Kristy Ann Phillips Nickles

All Rights Reserved

CLONING AND CHARACTERIZATION OF SUBTELOMERIC REGIONS OF

*KLUYVEROMYCES LACTIS*

by

KRISTY ANN PHILLIPS NICKLES

Major Professor: Michael McEachern

Committee: Sidney Kushner  
John McDonald

Electronic Version Approved:

Maureen Grasso  
Dean of the Graduate School  
The University of Georgia  
December 2003

DEDICATION

S.D.G.

## ACKNOWLEDGEMENTS

Many contributed to this work in a variety of ways. First, I'd like to thank Mike McEachern for the opportunity to work under his guidance. His flexibility and helpfulness certainly helped in making the environment more pleasant and the experience less stressful. Other students in the lab, Dana Underwood, Shobhana Natarajan, Shilpa Iyer, Chenoo Askree, and Cindy Vindman, have each earned my gratitude in several ways – from answering technical questions to being great friends. I thank the many undergraduates for contributing to the pleasant environment of the lab; I want to thank Will McRae and Ashley Chadha for specifically contributing to the work presented here.

All of the sequencing work of this study was performed by MGIF at UGA. The facility's willingness to work with me in the cost of this sequencing endeavor and in the occasional difficulty of getting good sequence have been invaluable as this project depended upon this sequence.

I also want to thank my committee, Sidney Kushner and John McDonald, for their service. Other labs of the Genetics Department offered assistance in lending various materials and equipment to the McEachern lab. Jessica Kissinger and Nandita Mullapudi gave advice regarding sequence analysis. I also want to thank the Kushner lab for supplying the plasmid on which my cloning vector was based. Such community within a department greatly benefits everyone involved.

My parents and brothers steadily offered support in their own ways during all my time in school. I especially want to acknowledge my husband, Lee Nickles, who was of vital support

during this time. His encouragement, help, and understanding have been key in keeping everything in balance.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS .....	v
CHAPTER	
1 INTRODUCTION AND LITERATURE REVIEW .....	1
2 SUBTELOMERIC REGIONS OF <i>KLUYVEROMYCES LACTIS</i> .....	27
3 CONCLUSIONS AND FUTURE DIRECTIONS.....	73
APPENDIX	
A pKN2 TRANSFORMANTS AND DERIVED PLASMIDS.....	80



## CHAPTER 1

### INTRODUCTION AND LITERATURE REVIEW

#### **Introduction**

The DNA of chromosome ends has two regions—telomeres and subtelomeric domains. While telomeres can usually be clearly defined (see below), the region called subterminal, subtelomeric, or telomere-associated sequence does not lend itself to such clear delineation. Subtelomeric regions in many organisms have a basic structure of homologous sequences at multiple chromosome ends, but there are no specific sequence similarities seen across species (Pryde et al., 1997). While the lack of conservation of subtelomeric DNA has previously led some to dismiss some or all of it as “junk” DNA, there are many examples, described below, that illustrate how this DNA provides advantages to particular species.

No definition of subtelomeric DNA appears well suited to all organisms. It is, therefore, useful to consider different ways of characterizing these regions. All of these, of course, are at least partly positional in nature; subtelomeric sequences must reside next to a telomere. Some possible descriptions of subtelomeric DNA are wholly positional in nature while others include functional characteristics of the sequences or repetitive structure. A usage of the word “subtelomeric” commonly seen in medical literature refers to any sequence relatively near the telomeres when viewing chromosomal karyotypes. This ad hoc description is not very useful for characterization at the molecular level. A simple molecular definition of subtelomeric DNA would be the region between the telomere and the closest gene (Figure 1.1A). This may prove to

be useful for organisms like *C. elegans* where the subtelomeric DNA is unique on each telomere (Wicky et al., 1996) but may be less useful in organisms with repetitive DNA near telomeres.

In principle, functional characteristics might also be used to define subtelomeric DNA. Sequences adjacent to telomeres that are somehow influenced by telomeres could be considered subtelomeric (Figure 1.1B). A well established example of this is transcriptional silencing of genes that are located near telomeres (Gottschling et al., 1990). Alternatively, sequences adjacent to telomeres that influence telomeric ends in *cis* could be defined as subtelomeric (Figure 1.1C). There is an example of this in *S. cerevisiae*, where subtelomeric sequences play a role in telomere length regulation (Craven and Petes, 1999). These definitions, while potentially useful, are impossible to apply in any but the best studied genetic model organisms. Even in *S. cerevisiae*, where they might best be used, the borders of the regions affected by telomeres or affecting telomere function have yet to be determined.

Perhaps the most useful definition of subtelomeric DNA, the definition that will usually be used in this work, is the region of a chromosome between telomeric DNA and the first single-copy gene (Wellinger and Sen, 1997) (Figure 1.1D). Repetitive DNA next to telomeres is one of the most common characteristics of chromosome ends across species. Furthermore, these repetitive sequences often have unusual characteristics. Thus, a more precise term for the sequences that will be discussed here would be ‘subtelomeric repetitive DNA’. As should become apparent by reading this work, subtelomeric DNA is more than just repetitive DNA that happens to be next to telomeres. Rather, it is repetitive DNA that has evolved special functions and characteristics because of its proximity to telomeres.

## **Telomere structure and function**

Telomeres are the DNA-protein complexes at the ends of linear chromosomes. In most eukaryotes, the DNA component consists of tandem copies of a short sequence (typically 5-8 bp) and is maintained by the enzyme telomerase (Blackburn, 1992). The other proteins that are part of this complex are recruited to these ends before and after new DNA synthesis to maintain functional telomeres. Telomeres preserve the chromosome structure in two ways; one, the telomerase prevents loss of non-telomeric sequence since a small amount of DNA from the chromosome tip is lost at each division due to the end-replication problem, and two, they “cap” chromosome ends so that they will not be recognized as broken ends in need of repair (Zakian, 1995; Shore, 1998; McEachern et al., 2000)

Typical telomeric DNA is G-rich on the strand running 5' to 3' toward the end and exists as an array of identical or variant tandem repeats. These arrays vary widely in length between different organisms. In the great majority of eukaryotes, telomeric repeats are relatively uniform in size at 5-8 bp in length (reviewed in (Wellinger and Sen, 1997)). However, in a group of budding yeasts that was studied, the repeats are 8-26 bp in length and usually comprise arrays of a few hundred bp per end (McEachern and Blackburn, 1994; Cohn et al., 1998). Most of the DNA in a telomere is double-stranded, but the very ends have 3' overhangs (Henderson and Blackburn, 1989; Wellinger et al., 1993; Makarov et al., 1997; Wright et al., 1997). Well-characterized structures observed that involve this overhang include the ‘t-loop’ structure (Griffith et al., 1999) and the G-quadruplex structure (Williamson, 1994). The t-loop is a large loop that has been observed at the ends of mammalian and protozoan telomeres *in vitro* and *in vivo* (Griffith et al., 1999; Munoz-Jordan et al., 2001). It is thought to be due to a D-loop forming between the 3' single strand end and a more internal region of the telomere (Figure

1.2A). The G-quadruplex is an intramolecular, four-stranded structure involving planar guanine quartets and has been seen with a variety of short telomeric repeat sequences *in vitro* (Figure 1.2B).

These DNA ends are maintained by telomerase, a ribonucleoprotein of which the catalytic subunit is a reverse transcriptase (Blackburn, 1992). It uses a portion of an RNA subunit as a template to synthesize DNA *de novo* onto chromosome ends (Greider and Blackburn, 1989). Its association with the telomeric end is dependent on interactions with telomeric proteins and on base pairing between the sequence at the very telomeric end and the template region of the RNA subunit. Studies *in vitro* and *in vivo* from a variety of species have indicated that it can act processively or not, depending on the species (Greider, 1991; Cohn and Blackburn, 1995).

A number of proteins besides telomerase are necessary to keep the telomeres fully functional (reviewed in McEachern et al., 2000). There are double-stranded and single-stranded DNA binding proteins among those known. Telomeric 3' single-stranded overhangs serve as the binding site for the single-stranded proteins and as the location of telomerase activity. Other proteins are associated via interactions with the DNA-bound proteins. Specific functions of the proteins binding to double-stranded DNA and those recruited by these proteins include negative regulation of telomere length, recruitment of other proteins, and telomere positioning within the nucleus (Palladino et al., 1993; Shore, 1997; Laroche et al., 1998). Functions of single-strand binding proteins include protection of the overhangs and recruitment of telomerase (Nugent et al., 1996).

A notable exception in telomere structure is that of *Drosophila melanogaster*. These ends consist of tandem arrays of non-LTR retrotransposons (Levis et al., 1993). Two types of

these elements are actively transposed to the ends, which prevents shortening in the same way that telomerase action prevents shortening. *D. melanogaster* lacks the typical sequence-dependent capping, as capping has been separated from the presence of the terminal retrotransposons (Biessmann et al., 1990). The protein HP1 (heterochromatin protein 1) has been shown to prevent telomere fusions in *D. melanogaster* with and without the presence of telomeric HeT-A and TART (Fanti et al., 1998). It has been suggested that HP1 is recruited by another protein not yet identified.

In humans, telomerase is active in germline and stem cells but typically inactive in other cells (Broccoli et al., 1995; Counter et al., 1995). It follows that in most somatic cells, over time, the ends shorten with each cell division (Harley et al., 1990). After about 50 generations in cell culture, normal somatic cells will reach replicative senescence due largely to effects of critically short telomeres (Harley et al., 1992; Hayflick, 1997; Karlseder et al., 2002; Wright and Shay, 2002). Replicative senescence is a permanent state in which vertebrate cells no longer divide despite lack of contact inhibition. This “life span” of human cells serves as one defense against the formation of immortal cancerous cells. Restoration of telomere maintenance is one means of bypassing replicative senescence and one of the key properties of immortal human cell lines and cancers. In most human cancers, telomeres are maintained by telomerase that is thought to have been transcriptionally activated during carcinogenesis (Kim et al., 1994; Shay and Bacchetti, 1997). In a minority of human cancers, telomerase is absent and telomeres are thought to be maintained by recombination (Bryan et al., 1997; Dunham et al., 2000).

### ***Saccharomyces cerevisiae* subtelomeric regions**

The subtelomeric regions of *S. cerevisiae* are the best characterized of any organism to date (Figure 1.3) (Pryde et al., 1997). Internal to the ~350 bp of telomeric repeats, the Y' elements are commonly located. These elements are highly conserved (commonly 99% within strains and 98% between strains) and are restricted to *S. cerevisiae* and its close relative *S. paradoxus*. The two most common types are classified as Y' long (6.7 kb) and Y' short (5.4 kb); this size difference results from insertions and deletions. These elements are present in 1-4 tandem copies per telomere and are present at a number of different chromosome ends that vary from strain to strain; in S288C, the fully sequenced strain, Y' elements are located at 17 of the 32 ends (Louis, 1995).

These Y' elements contain two open reading frames (ORFs). One of these ORFs has been shown to be a DNA helicase by sequence homology (Louis and Haber, 1992) and biochemical analysis (Yamada et al., 1998). There is little expression of either ORF during vegetative growth, but there is more during meiosis. There are also repetitive sequences in these elements, chiefly in the form of minisatellites (Horowitz and Haber, 1984) and repeats that have degenerate telomeric repeats. Y' elements also contain autonomously replicating sequences (ARS) (Louis and Haber, 1992).

The X element is the other major subtelomeric element in *S. cerevisiae*. It is a mosaic element that is more variable than the Y' elements. There is a 475-bp consensus, the 'core' X, which contains an ARS and Abf1p (ARS binding factor) binding site. This core element is present in some form at all 32 ends in *S. cerevisiae* at positions internal to the Y' element (Figure 1.3) (Louis et al., 1994). The variable outer portion of this element, which is not always present, is composed of families of tandem degenerate repeats. The four common characterized groups

of repetitive elements are designated subtelomeric repeat A (STR-A), STR-B, STR-C, and STR-D. These elements range 30-150 bp in length and are fairly well conserved (around 90% identical in pairwise comparisons). Presence and copy-number of these repeats vary among chromosome ends, resulting in the X element's mosaic nature.

The *SUC* and *RTM* gene families are associated with the X element (Carlson et al., 1985; Ness and Aigle, 1995). The *SUC* genes encode copies of an invertase involved in sucrose utilization ( $\beta$ -fructofuranosidase), and the *RTM* gene encodes a protein that, when overexpressed, confers resistance to molasses. The specific function of *RTM* proteins has not been identified.

Internal to the X element, there is varying homology among the ends extending for 2-30 kb in S288C. The larger blocks of homology that make up the proximal domains exist in ten groups with common blocks of genes being shared between two to three ends (<http://www.le.ac.uk/genetics/ejl12/research/telostruc/ClustersSmall.html>, Ed Louis, personal communication). This is where other gene families such as *MAL*, *MEL* and *FLO* are often found. *MAL* genes are necessary for maltose catabolism ( $\alpha$ -glucosidase), and *MEL* genes are necessary in melibiose catabolism ( $\alpha$ -galactosidase) (Pryde and Louis, 1997). *FLO* genes encode copies of the surface protein flocculin necessary for flocculation (cell surface adhesion) (Teunissen and Steensma, 1995). The presence, location, and copy number of these genes vary greatly from strain to strain. Typically, a given strain will have either *SUC* or *MEL* genes but not both, suggesting there is competition between these two families (Naumov et al., 1996). Southern data show that *SUC* genes are found in bakers' and molasses alcohol distillers' strains at several positions and higher copy number than other industrial strains (Codon et al., 1998). Most natural strains contain more than one *MAL* locus in their genomes (Naumov et al., 1994). These differences seen across strains appear to correlate to the industrial use and environment of the

strains, so it seems that this region has been used by yeast populations in adapting to the specialized environments of industry (Codon et al., 1998).

One or more copies of the yeast retrotransposon Ty5 is also found in the subtelomeric regions of *S. cerevisiae* and *S. paradoxus*. This LTR-transposon is variable in copy number (0-8) and location within and between strains. The endogenous Ty5 elements found in *S. cerevisiae* are not transposition competent (Zou et al., 1996b). An active element from *S. paradoxus* has been shown to have a preference for insertion near the X element of *S. cerevisiae* (Zou et al., 1995; Zou et al., 1996a).

### **Subtelomeric regions of other organisms**

A common description of subtelomeric sequences is that they can be divided into two general zones or domains: a distal domain, adjacent to the telomere, that is homologous among a relatively large number of telomeres and a proximal domain that is homologous among fewer telomeres. Another feature seen in different species is variability in the dosage and location of some subtelomeric genes such as is seen in the *SUC*, *MAL*, and *MEL* gene families of industrial yeasts discussed above.

Human subtelomeric domains share structural similarities to those of *S. cerevisiae* (Figure 1.3) (Flint et al., 1997). These subtelomeric regions are commonly thought to extend for 100-300 kb (Brown et al., 1990; Martin et al., 2002). Comparison of 60 kb of sequence from 4q to the other ends showed that 17 other ends have non-contiguous similarities of >90% in the distal 15 kb of non-telomeric sequence of that arm. In the 45 kb proximal to that region, only four other arms showed matches (Flint et al., 1997). Human subtelomeric regions can also contain areas of degenerate telomeric repeats near the telomere that appear to separate the distal



region from the proximal region. As in *S. cerevisiae*, duplicated gene families also exist in this proximal region. For example, copies of a characterized block including three ORFs similar to the olfactory receptor genes have been seen on at least 14 different ends (Mefford et al., 2001) and vary in number, location, and transcriptional activity (Trask et al., 1998; Linardopoulou et al., 2001)

The eukaryotic parasitic protozoans *Plasmodium falciparum* and *Trypanosoma brucei*, also have a general subtelomeric structure reminiscent of yeast and humans (Figure 1.3). Along with repetitive elements that are shared among ends, these organisms have their surface antigen genes in this region (Rubio et al., 1996; Borst and Ulbert, 2001). This is discussed further below.

In *Drosophila melanogaster*, chromosome ends are maintained by transpositions of non-LTR retrotransposons Het-A and TART instead of telomerase-mediated elongation (Levis et al., 1993). Despite this difference in telomere function and maintenance from most eukaryotes, its subtelomeric structure is similar to that of other eukaryotes that use telomerase (Figure 1.3) in that they contain several kilobases of tandem repeats (Mason and Biessmann, 1995). These repeats are not identical and can have subrepeats within them. Some ends are known to cross-hybridize with each other by *in situ* hybridization indicating a degree of homology among these ends.

The generally conserved structure for subtelomeric regions testifies to its important role in genome organization. However, not all organisms fit into this general description.

*Caenorhabditis elegans* does not appear to have the traditional type of subtelomeric region.

There are repetitive sequences next to telomeres, but there is no homology between any of the ends (Wicky et al., 1996).

## **Functions and properties of subtelomeric regions**

In studying subtelomeric regions of many organisms, the plasticity found here has become a major theme for investigation. Plasticity can manifest itself in a variety of ways. In such regions of a genome, one would expect to observe trends such as higher rates of recombination and rearrangements which would not be practical in “less plastic” regions. Examples of such observations are described below.

The general structure of subtelomeric homology between many telomeres, which emerges from many studies, has been suggested to confer an ability to use recombination at much higher rates than is typical for the rest of the chromosome (Pryde et al., 1997). In the malaria parasite *P. falciparum*, genes for surface antigens are found in subtelomeric regions; antigenic variations of PfEMP-1 (erythrocyte membrane protein 1), the major variable parasite protein, are encoded by the *var* genes. There are 50 to 150 members of this family, most of which are located near the ends of the chromosomes. A typical haploid genome contains 50 members. It has been shown that ectopic recombination, homologous recombination between non-homologous chromosomes, can occur at a high rate during meiosis in *P. falciparum* to create new *var* genes, which would enable efficient evasion of hosts' immune systems (Freitas-Junior et al., 2000). This rate is higher than ectopic rates between *var* genes not located near telomeres. Subtelomeric regions also appear to be necessary in forming the telomeric clusters during which these ectopic exchanges take place (Figueiredo et al., 2002). Clearly, the organization and interaction of subtelomeric regions is advantageous to this organism.

Recombination in subtelomeric regions has also been shown to play a role in telomere maintenance in cells that lack a functional telomerase (McEachern and Blackburn, 1996). The effects of loss of telomere function have been studied by deleting genes encoding subunits of

telomerase including the telomerase RNA gene. In the budding yeasts *S. cerevisiae* and *K. lactis*, deletion of telomerase leads to telomere shortening and growth senescence (Lundblad and Szostak, 1989; McEachern and Blackburn, 1996). However, surviving telomerase-deficient cells will restore substantially elongated telomeres (Lundblad and Blackburn, 1993). This act of survival has been shown to employ homologous recombination as a means of restoring telomere length. Recombination events that would restore telomeres can occur within telomeres themselves or in the subtelomeric region.

In *S. cerevisiae*, there are two types of post-senescence survivors seen in telomerase deletion strains. They can be classified according to the behavior of the Y' elements and of the telomeric repeat arrays. In Type 1 survivors, Y' elements have undergone amplification while telomeres remain short; in Type 2 survivors, Y' amplification is not seen and instead the telomeric repeat arrays have been elongated (Teng and Zakian, 1999).

Genomic DNA from *K. lactis* telomerase-RNA template (*TER1*) mutants with short telomeres that were followed over many generations shows that changes commonly occur in the subtelomeric regions of the chromosomes (McEachern and Iyer, 2001). With little subtelomeric sequence information, these changes could only be seen as altered restriction patterns of *K. lactis* genomic DNA.

The previous examples deal with relatively minor or routine changes of chromosome ends. Perhaps the most dramatic change a chromosome end can undergo is a complete deletion of its telomeric repeat tract. It is likely that subtelomeric sequences would be intimately involved in the repair of this end, which would act as a double strand break (DSB). DSBs have three possible mechanisms of repair. One is simply adding telomeric repeats to the broken end, which is known as chromosome healing. This is regularly seen in organisms such as

*Tetrahymena thermophila* (Fan and Yao, 1996), *Euplotes crassus* (Roth and Prescott, 1985), *Paramecium tetraurelia* (Forney and Blackburn, 1988), *Cryptococcus neoformans* (Edman, 1992), and *Plasmodium falciparum* (Pologe and Ravetch, 1988). This method of repairing a DNA end lacking telomeric repeats is uncommon in other organisms. In *S. cerevisiae*, one study reports that in a strain deficient in homologous recombination, experimentally generated DSBs were healed by new telomere addition 1% of the time (Kramer and Haber, 1993). Also, in humans chromosome healing has been observed connected with the subtelomeric truncation that causes the genetic disease  $\alpha$ -thalassemia (Wilkie et al., 1990; Flint et al., 1996).

Usually, DSB repair is accomplished by either nonhomologous end-joining (NHEJ) or by recombinational repair that uses a sister chromatid or a homolog (Lundblad, 2000). The more common pathway in yeast, recombinational repair, can affect subtelomeric structure by eventually homogenizing regions near all of the ends (McEachern and Iyer, 2001).

Another interesting property of subtelomeric regions that has been studied is telomere position effect (TPE), which involves the silencing of genes in this region of the genome. In *S. cerevisiae* and *K. lactis*, reporter genes that are part of cassettes replacing sequence next to the telomeres are commonly silenced (Gottschling et al., 1990; Gurevich et al., 2003). In these cases, the closer to the telomere the reporter gene is, the higher the rate of repression (Renauld et al., 1993). However, a larger range of repression is seen in natural telomeres (Pryde and Louis, 1999). This difference among ends is tentatively attributed to the mosaic nature of the subtelomeric regions (Tham and Zakian, 2002). Humans also exhibit TPE that appears to be affected by telomere length. Genes near the telomere are expressed at a tenth of the normal level, and expression can be restored with treatment of a histone deacetylase inhibitor (Baur et

al., 2001). Human TPE has implications worth notice since human telomeres shorten with replicative age and possibly could have effects on expression of genes near telomeres.

### **Potential uses of subtelomeric sequences**

Subtelomeric sequences can serve as valuable tools. It is often useful to identify specific chromosome ends. For a particular strain, polymorphisms and other chromosome-specific features can be identified and used to follow chromosomes over time. This can offer a way of identifying subtelomeric rearrangements or tracing these rearrangements back to their ancestral sequence. The latter of these two would prove useful in examining the adaptations that are facilitated by subtelomeric structure.

In humans, the high level of change in subtelomeric regions can have deleterious effects. The etiology of a portion of cases of idiopathic mental retardation can be traced to chromosomal rearrangements in subtelomeric sequences (Wilkie, 1993; Flint et al., 1995). In familial cases, rearrangements that segregate with phenotypes are observed. Groups are moving to establish the clinical use of subtelomeric probes in these cases and other disorders associated with abnormalities at chromosome ends (Knight and Flint, 2000).

Subtelomeric sequences have been known to affect and be affected by telomeres; therefore, in organisms of telomere/telomerase studies, detailed knowledge of these regions would be useful in exploring this relationship. For example, telomere studies in *K. lactis* often involve recombinational telomere elongation, which appears to bring about alterations of subtelomeric sequences. This study reports efforts to characterize the subtelomeric regions of *K. lactis* that included searches for genes, analysis of homology between different chromosome ends, and searches for any *K. lactis*-specific subtelomeric motifs.

## References

- Baur, J. A., Zou, Y., Shay, J. W. and Wright, W. E. (2001). Telomere position effect in human cells. *Science* **292**: 2075-7.
- Biessmann, H., Carter, S. B. and Mason, J. M. (1990). Chromosome ends in *Drosophila* without telomeric DNA sequences. *Proc Natl Acad Sci U S A* **87**: 1758-61.
- Blackburn, E. H. (1992). Telomerases. *Annu Rev Biochem* **61**: 113-29.
- Borst, P. and Ulbert, S. (2001). Control of VSG gene expression sites. *Mol Biochem Parasitol* **114**: 17-27.
- Broccoli, D., Young, J. W. and de Lange, T. (1995). Telomerase activity in normal and malignant hematopoietic cells. *Proc Natl Acad Sci U S A* **92**: 9082-6.
- Brown, W. R., MacKinnon, P. J., Villasante, A., Spurr, N., Buckle, V. J. and Dobson, M. J. (1990). Structure and polymorphism of human telomere-associated DNA. *Cell* **63**: 119-32.
- Bryan, T. M., Englezou, A., Dalla-Pozza, L., Dunham, M. A. and Reddel, R. R. (1997). Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat Med* **3**: 1271-4.
- Carlson, M., Celenza, J. L. and Eng, F. J. (1985). Evolution of the dispersed SUC gene family of *Saccharomyces* by rearrangements of chromosome telomeres. *Mol Cell Biol* **5**: 2894-902.
- Codon, A. C., Benitez, T. and Korhola, M. (1998). Chromosomal polymorphism and adaptation to specific industrial environments of *Saccharomyces* strains. *Appl Microbiol Biotechnol* **49**: 154-63.
- Cohn, M. and Blackburn, E. H. (1995). Telomerase in yeast. *Science* **269**: 396-400.

- Cohn, M., McEachern, M. J. and Blackburn, E. H. (1998). Telomeric sequence diversity within the genus *Saccharomyces*. *Curr Genet* **33**: 83-91.
- Counter, C. M., Gupta, J., Harley, C. B., Leber, B. and Bacchetti, S. (1995). Telomerase activity in normal leukocytes and in hematologic malignancies. *Blood* **85**: 2315-20.
- Craven, R. J. and Petes, T. D. (1999). Dependence of the regulation of telomere length on the type of subtelomeric repeat in the yeast *Saccharomyces cerevisiae*. *Genetics* **152**: 1531-41.
- Dunham, M. A., Neumann, A. A., Fasching, C. L. and Reddel, R. R. (2000). Telomere maintenance by recombination in human cells. *Nat Genet* **26**: 447-50.
- Edman, J. C. (1992). Isolation of telomere-like sequences from *Cryptococcus neoformans* and their use in high-efficiency transformation. *Mol Cell Biol* **12**: 2777-83.
- Fan, Q. and Yao, M. (1996). New telomere formation coupled with site-specific chromosome breakage in *Tetrahymena thermophila*. *Mol Cell Biol* **16**: 1267-74.
- Fanti, L., Giovinazzo, G., Berloco, M. and Pimpinelli, S. (1998). The heterochromatin protein 1 prevents telomere fusions in *Drosophila*. *Mol Cell* **2**: 527-38.
- Figueiredo, L. M., Freitas-Junior, L. H., Bottius, E., Olivo-Marin, J. C. and Scherf, A. (2002). A central role for *Plasmodium falciparum* subtelomeric regions in spatial positioning and telomere length regulation. *EMBO J* **21**: 815-24.
- Flint, J., Bates, G. P., Clark, K., Dorman, A., Willingham, D., Roe, B. A., Micklem, G., Higgs, D. R. and Louis, E. J. (1997). Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Hum Mol Genet* **6**: 1305-13.
- Flint, J., Rochette, J., Craddock, C. F., Dode, C., Vignes, B., Horsley, S. W., Kearney, L., Buckle, V. J., Ayyub, H. and Higgs, D. R. (1996). Chromosomal stabilisation by a

- subtelomeric rearrangement involving two closely related Alu elements. *Hum Mol Genet* **5**: 1163-9.
- Flint, J., Wilkie, A. O., Buckle, V. J., Winter, R. M., Holland, A. J. and McDermid, H. E. (1995). The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation. *Nat Genet* **9**: 132-40.
- Forney, J. D. and Blackburn, E. H. (1988). Developmentally controlled telomere addition in wild-type and mutant paramecia. *Mol Cell Biol* **8**: 251-8.
- Freitas-Junior, L. H., Bottius, E., Pirrit, L. A., Deitsch, K. W., Scheidig, C., Guinet, F., Nehrass, U., Wellems, T. E. and Scherf, A. (2000). Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* **407**: 1018-22.
- Gottschling, D. E., Aparicio, O. M., Billington, B. L. and Zakian, V. A. (1990). Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell* **63**: 751-62.
- Greider, C. W. (1991). Telomerase is processive. *Mol Cell Biol* **11**: 4572-80.
- Greider, C. W. and Blackburn, E. H. (1989). A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature* **337**: 331-7.
- Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H. and de Lange, T. (1999). Mammalian telomeres end in a large duplex loop. *Cell* **97**: 503-14.
- Gurevich, R., Smolikov, S., Maddar, H. and Krauskopf, A. (2003). Mutant telomeres inhibit transcriptional silencing at native telomeres of the yeast *Kluyveromyces lactis*. *Mol Genet Genomics* **268**: 729-38.
- Harley, C. B., Futcher, A. B. and Greider, C. W. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature* **345**: 458-60.



- Harley, C. B., Vaziri, H., Counter, C. M. and Allsopp, R. C. (1992). The telomere hypothesis of cellular aging. *Exp Gerontol* **27**: 375-82.
- Hayflick, L. (1997). Mortality and immortality at the cellular level. A review. *Biochemistry (Mosc)* **62**: 1180-90.
- Henderson, E. R. and Blackburn, E. H. (1989). An overhanging 3' terminus is a conserved feature of telomeres. *Mol Cell Biol* **9**: 345-8.
- Horowitz, H. and Haber, J. E. (1984). Subtelomeric regions of yeast chromosomes contain a 36 base-pair tandemly repeated sequence. *Nucleic Acids Res* **12**: 7105-21.
- Karlseder, J., Smogorzewska, A. and de Lange, T. (2002). Senescence induced by altered telomere state, not telomere loss. *Science* **295**: 2446-9.
- Kim, N. W., Piatyszek, M. A., Prowse, K. R., Harley, C. B., West, M. D., Ho, P. L., Coviello, G. M., Wright, W. E., Weinrich, S. L. and Shay, J. W. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**: 2011-5.
- Knight, S. J. and Flint, J. (2000). Perfect endings: a review of subtelomeric probes and their use in clinical diagnosis. *J Med Genet* **37**: 401-409.
- Kramer, K. M. and Haber, J. E. (1993). New telomeres in yeast are initiated with a highly selected subset of TG1-3 repeats. *Genes Dev* **7**: 2345-56.
- Laroche, T., Martin, S. G., Gotta, M., Gorham, H. C., Pryde, F. E., Louis, E. J. and Gasser, S. M. (1998). Mutation of yeast Ku genes disrupts the subnuclear organization of telomeres. *Curr Biol* **8**: 653-6.
- Levis, R. W., Ganesan, R., Houtchens, K., Tolar, L. A. and Sheen, F. M. (1993). Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* **75**: 1083-93.

- Linardopoulou, E., Mefford, H. C., Nguyen, O., Friedman, C., van den Engh, G., Farwell, D. G., Coltrera, M. and Trask, B. J. (2001). Transcriptional activity of multiple copies of a subtelomerically located olfactory receptor gene that is polymorphic in number and location. *Hum Mol Genet* **10**: 2373-83.
- Louis, E. J. (1995). The chromosome ends of *Saccharomyces cerevisiae*. *Yeast* **11**: 1553-73.
- Louis, E. J. and Haber, J. E. (1992). The structure and evolution of subtelomeric Y' repeats in *Saccharomyces cerevisiae*. *Genetics* **131**: 559-74.
- Louis, E. J., Naumova, E. S., Lee, A., Naumov, G. and Haber, J. E. (1994). The chromosome end in yeast: its mosaic nature and influence on recombinational dynamics. *Genetics* **136**: 789-802.
- Lundblad, V. (2000). DNA ends: maintenance of chromosome termini versus repair of double strand breaks. *Mutat Res* **451**: 227-40.
- Lundblad, V. and Blackburn, E. H. (1993). An alternative pathway for yeast telomere maintenance rescues *est1*- senescence. *Cell* **73**: 347-60.
- Lundblad, V. and Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* **57**: 633-43.
- Makarov, V. L., Hirose, Y. and Langmore, J. P. (1997). Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. *Cell* **88**: 657-66.
- Martin, C. L., Wong, A., Gross, A., Chung, J., Fantes, J. A. and Ledbetter, D. H. (2002). The evolutionary origin of human subtelomeric homologies--or where the ends begin. *Am J Hum Genet* **70**: 972-84.

- Mason, J. M. and Biessmann, H. (1995). The unusual telomeres of *Drosophila*. *Trends Genet* **11**: 58-62.
- McEachern, M. J. and Blackburn, E. H. (1994). A conserved sequence motif within the exceptionally diverse telomeric sequences of budding yeasts. *Proc Natl Acad Sci U S A* **91**: 3453-7.
- McEachern, M. J. and Blackburn, E. H. (1996). Cap-prevented recombination between terminal telomeric repeat arrays (telomere CPR) maintains telomeres in *Kluyveromyces lactis* lacking telomerase. *Genes Dev* **10**: 1822-34.
- McEachern, M. J. and Iyer, S. (2001). Short telomeres in yeast are highly recombinogenic. *Mol Cell* **7**: 695-704.
- McEachern, M. J., Krauskopf, A. and Blackburn, E. H. (2000). Telomeres and Their Control. *Annu Rev Genet* **34**: 331-358.
- Mefford, H. C., Linardopoulou, E., Coil, D., van den Engh, G. and Trask, B. J. (2001). Comparative sequencing of a multicopy subtelomeric region containing olfactory receptor genes reveals multiple interactions between non-homologous chromosomes. *Hum Mol Genet* **10**: 2363-72.
- Munoz-Jordan, J. L., Cross, G. A., de Lange, T. and Griffith, J. D. (2001). t-loops at trypanosome telomeres. *EMBO J* **20**: 579-88.
- Naumov, G. I., Naumova, E. S. and Michels, C. A. (1994). Genetic variation of the repeated *MAL* loci in natural populations of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*. *Genetics* **136**: 803-12.
- Naumov, G. I., Naumova, E. S., Sancho, E. D. and Korhola, M. P. (1996). Polymeric *SUC* genes in natural populations of *Saccharomyces cerevisiae*. *FEMS Microbiol Lett* **135**: 31-5.

- Ness, F. and Aigle, M. (1995). *RTM1*: a member of a new family of telomeric repeated genes in yeast. *Genetics* **140**: 945-56.
- Nugent, C. I., Hughes, T. R., Lue, N. F. and Lundblad, V. (1996). Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance. *Science* **274**: 249-52.
- Palladino, F., Laroche, T., Gilson, E., Axelrod, A., Pillus, L. and Gasser, S. M. (1993). *SIR3* and *SIR4* proteins are required for the positioning and integrity of yeast telomeres. *Cell* **75**: 543-55.
- Pologe, L. G. and Ravetch, J. V. (1988). Large deletions result from breakage and healing of *P. falciparum* chromosomes. *Cell* **55**: 869-74.
- Pryde, F. E., Gorham, H. C. and Louis, E. J. (1997). Chromosome ends: all the same under their caps. *Curr Opin Genet Dev* **7**: 822-8.
- Pryde, F. E. and Louis, E. J. (1997). *Saccharomyces cerevisiae* telomeres. A review. *Biochemistry* **62**: 1232-41.
- Pryde, F. E. and Louis, E. J. (1999). Limitations of silencing at native yeast telomeres. *EMBO J* **18**: 2538-2550.
- Renauld, H., Aparicio, O. M., Zierath, P. D., Billington, B. L., Chhablani, S. K. and Gottschling, D. E. (1993). Silent domains are assembled continuously from the telomere and are defined by promoter distance and strength, and by *SIR3* dosage. *Genes Dev* **7**: 1133-45.
- Roth, M. and Prescott, D. M. (1985). DNA intermediates and telomere addition during genome reorganization in *Euplotes crassus*. *Cell* **41**: 411-7.

- Rubio, J. P., Thompson, J. K. and Cowman, A. F. (1996). The var genes of *Plasmodium falciparum* are located in the subtelomeric region of most chromosomes. *EMBO J* **15**: 4069-77.
- Shay, J. W. and Bacchetti, S. (1997). A survey of telomerase activity in human cancer. *Eur J Cancer* **33**: 787-91.
- Shore, D. (1997). Telomere length regulation: getting the measure of chromosome ends. *Biol Chem* **378**: 591-7.
- Shore, D. (1998). Telomeres--unsticky ends. *Science* **281**: 1818-9.
- Teng, S. C. and Zakian, V. A. (1999). Telomere-telomere recombination is an efficient bypass pathway for telomere maintenance in *Saccharomyces cerevisiae*. *Mol Cell Biol* **19**: 8083-93.
- Teunissen, A. W. and Steensma, H. Y. (1995). Review: the dominant flocculation genes of *Saccharomyces cerevisiae* constitute a new subtelomeric gene family. *Yeast* **11**: 1001-13.
- Tham, W. H. and Zakian, V. A. (2002). Transcriptional silencing at *Saccharomyces* telomeres: implications for other organisms. *Oncogene* **21**: 512-21.
- Trask, B. J., Friedman, C., Martin-Gallardo, A., Rowen, L., Akinbami, C., Blankenship, J., Collins, C., Giorgi, D., Iadonato, S., Johnson, F., Kuo, W. L., Massa, H., Morrish, T., Naylor, S., Nguyen, O. T., Rouquier, S., Smith, T., Wong, D. J., Youngblom, J. and van den Engh, G. (1998). Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum Mol Genet* **7**: 13-26.
- Wellinger, R. J. and Sen, D. (1997). The DNA structures at the ends of eukaryotic chromosomes. *Eur J Cancer* **33**: 735-49.

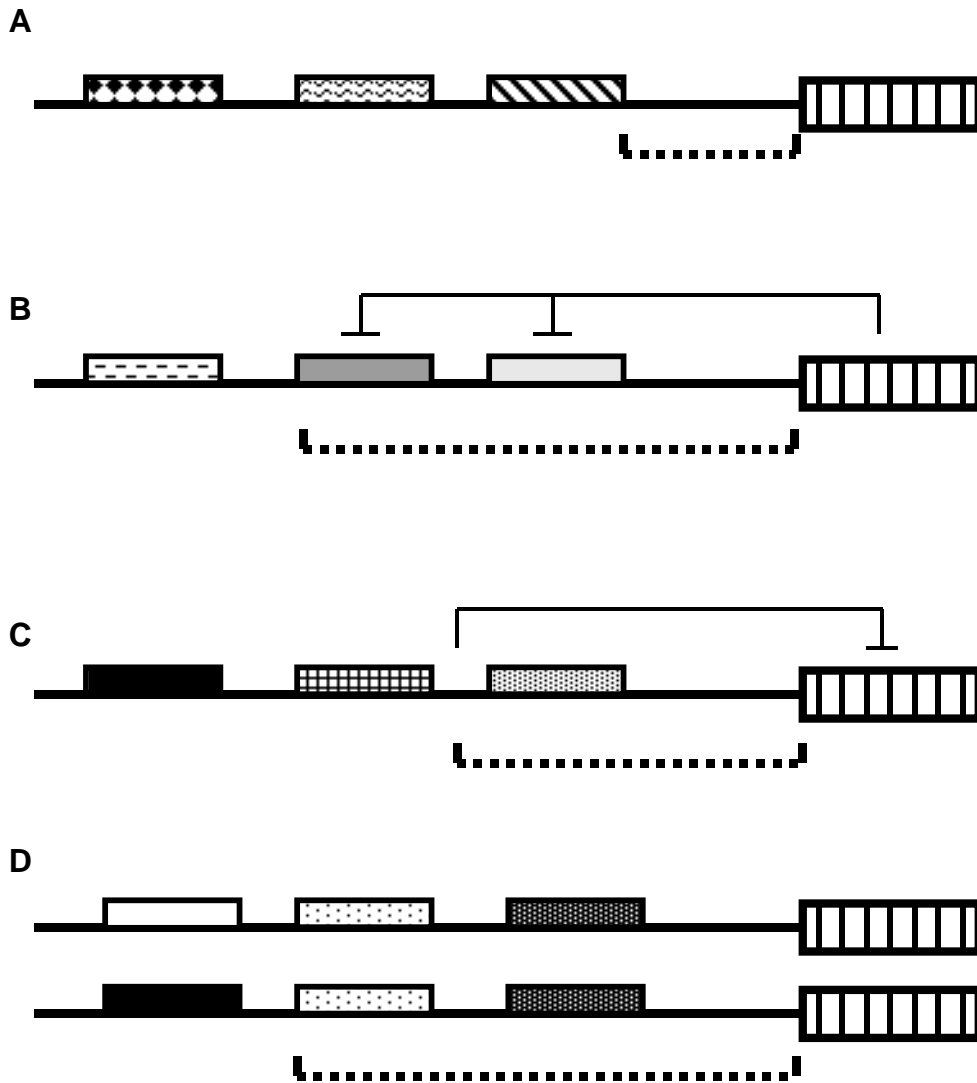
- Wellinger, R. J., Wolf, A. J. and Zakian, V. A. (1993). Saccharomyces telomeres acquire single-strand TG1-3 tails late in S phase. *Cell* **72**: 51-60.
- Wicky, C., Villeneuve, A. M., Lauper, N., Codourey, L., Tobler, H. and Muller, F. (1996). Telomeric repeats (TTAGGC)<sub>n</sub> are sufficient for chromosome capping function in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* **93**: 8983-8.
- Wilkie, A. O. (1993). Detection of cryptic chromosomal abnormalities in unexplained mental retardation: a general strategy using hypervariable subtelomeric DNA polymorphisms. *Am J Hum Genet* **53**: 688-701.
- Wilkie, A. O., Lamb, J., Harris, P. C., Finney, R. D. and Higgs, D. R. (1990). A truncated human chromosome 16 associated with alpha thalassaemia is stabilized by addition of telomeric repeat (TTAGGG)<sub>n</sub>. *Nature* **346**: 868-71.
- Williamson, J. R. (1994). G-quartet structures in telomeric DNA. *Annu Rev Biophys Biomol Struct* **23**: 703-30.
- Wright, W. E. and Shay, J. W. (2002). Historical claims and current interpretations of replicative aging. *Nat Biotechnol* **20**: 682-8.
- Wright, W. E., Tesmer, V. M., Huffman, K. E., Levene, S. D. and Shay, J. W. (1997). Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes Dev* **11**: 2801-9.
- Yamada, M., Hayatsu, N., Matsuura, A. and Ishikawa, F. (1998). Y'-Help1, a DNA helicase encoded by the yeast subtelomeric Y' element, is induced in survivors defective for telomerase. *J Biol Chem* **273**: 33360-6.
- Zakian, V. A. (1995). Telomeres: beginning to understand the end. *Science* **270**: 1601-7.

Zou, S., Ke, N., Kim, J. M. and Voytas, D. F. (1996a). The *Saccharomyces* retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci.

*Genes Dev* **10**: 634-45.

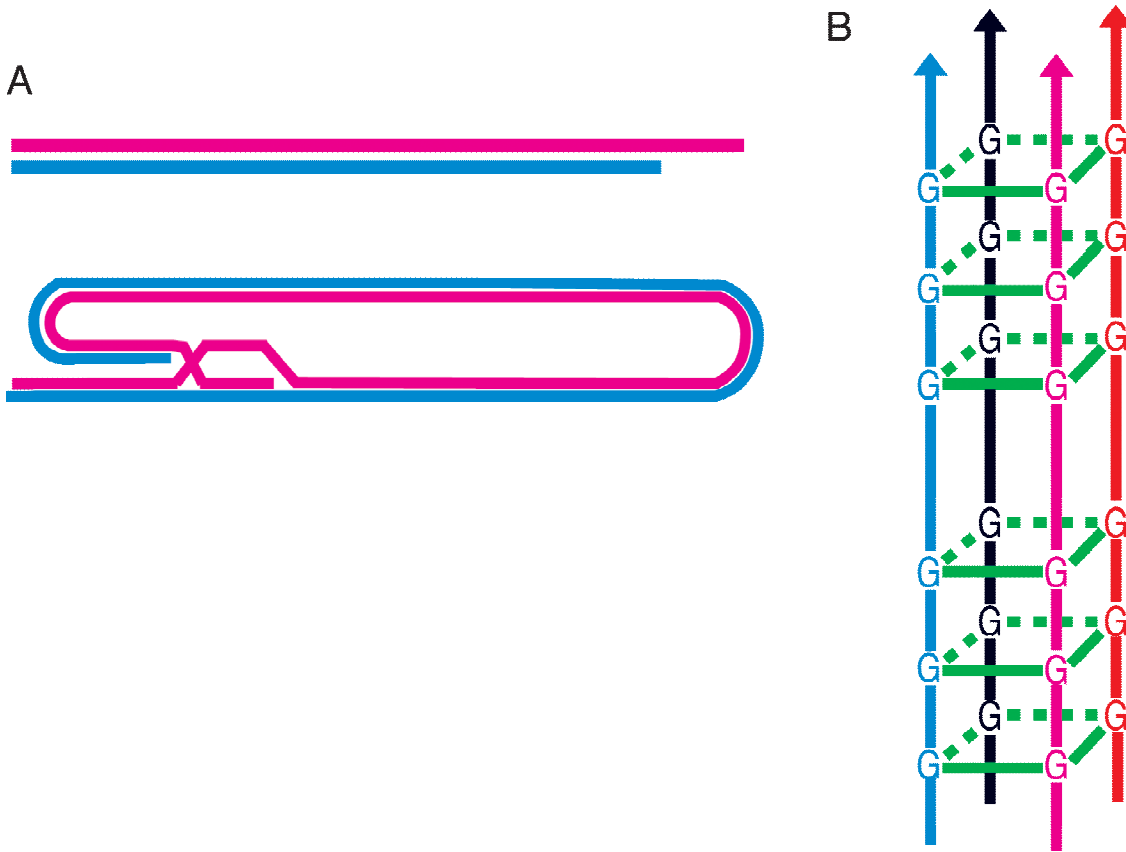
Zou, S., Kim, J. M. and Voytas, D. F. (1996b). The *Saccharomyces* retrotransposon Ty5 influences the organization of chromosome ends. *Nucleic Acids Res* **24**: 4825-31.

Zou, S., Wright, D. A. and Voytas, D. F. (1995). The *Saccharomyces* Ty5 retrotransposon family is associated with origins of DNA replication at the telomeres and the silent mating locus HMR. *Proc Natl Acad Sci U S A* **92**: 920-4.

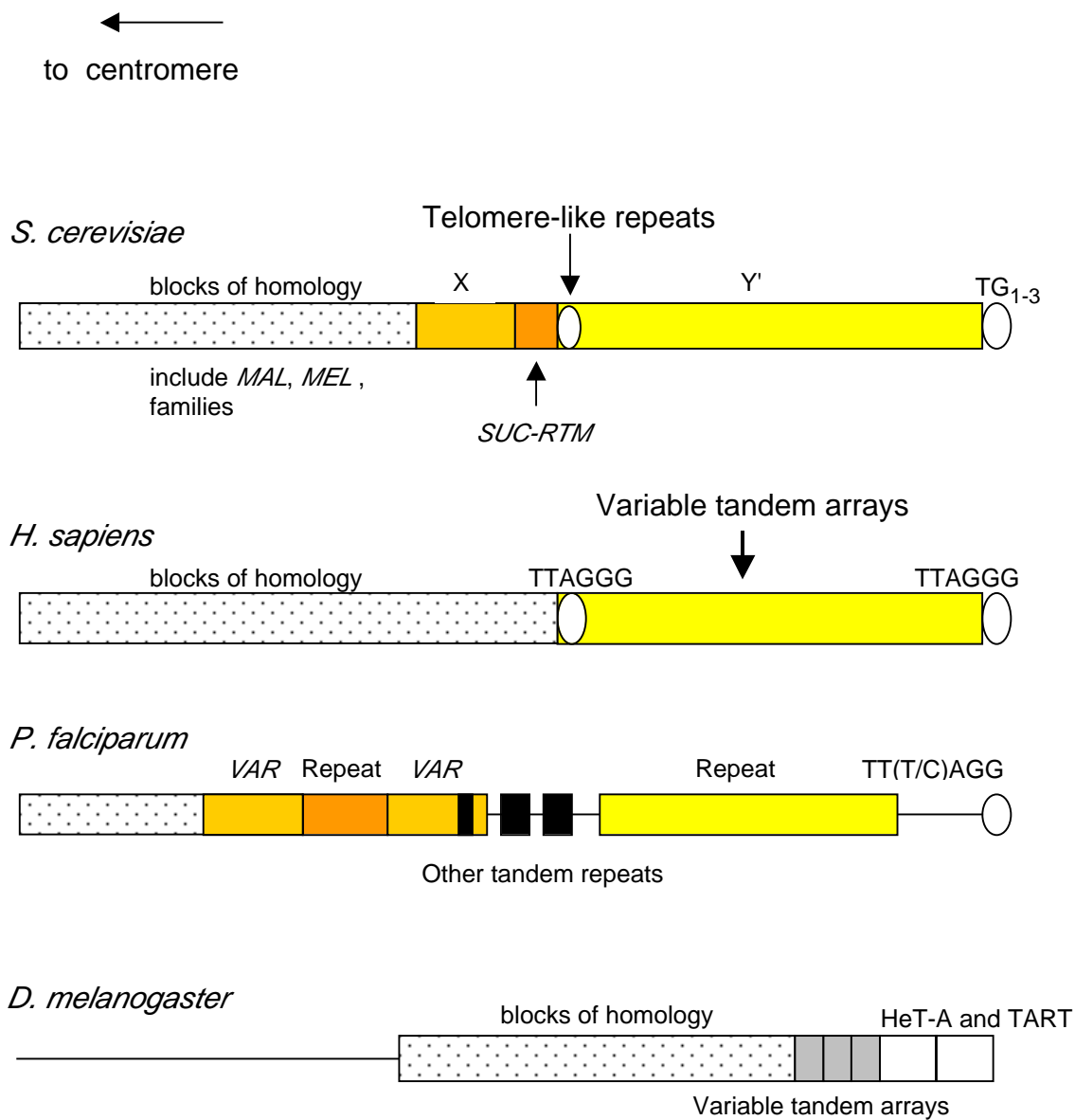


**Figure 1.1: Possible definitions of subtelomeric regions.** Blocks on the right represent telomeric repeats, and all other blocks represent genes – each pattern represents a different gene. Dashed lines in each panel show the region that would be the subtelomere for that definition. **A)** region internal to the telomere with no genes. **B)** adjacent sequence that is influenced by telomeres. **C)** adjacent sequence that influences the telomere. **D)** region between telomeric repeats and first single copy gene.





**Figure 1.2: Examples of structures that involve telomeric G-overhang.** **A)** Overhang shown straight and as part of a duplex loop known as a t-loop. The pink strand runs 5' to 3', and the blue runs 3' to 5'. **B)** Example of a G-quadruplex of four different strands is shown. Each color represents a different molecule in the structure, and solid and dashed green lines represent hydrogen bonds.



**Figure 1.3: Subtelomeric regions discussed throughout introduction.** Adapted from (Pryde et al., 1997). Circles throughout represent respective telomeres or telomere-like repeats as labeled; in the *D. melanogaster* panel the white rectangles represent the retrotransposon ends. General positions of selected subtelomeric genes and repeats are labeled. Solid lines show unique sequence or uncharacterized segments.

## CHAPTER 2

### SUBTELOMERIC REGIONS OF *KLUYVEROMYCES LACTIS*<sup>1</sup>

---

<sup>1</sup> Nickles, K.P. and M.J. McEachern. To be submitted to *Yeast*.

## **Abstract**

Telomeres are the specialized structures at the ends of eukaryotic chromosomes and are composed of short T/G-rich DNA repeats and the proteins that interact with them. Internal to telomeres are subtelomeric regions that are species-specific and often repetitive. The yeast *Kluyveromyces lactis* has telomeric tracts of 10-20 copies of a 25-bp repeat, but the subtelomeric regions have not previously been characterized in detail. Here we have cloned and characterized subtelomeric regions from ten of the twelve chromosome ends. The amount of sequence examined ranged 0.7-10 kb for each subtelomeric region. We have identified a *K. lactis* subtelomeric element, the R element, which has a purine/pyrimidine strand bias and extends for about 2 kb. Internal to the R element, we found extensive homology among half of the chromosome ends reported here. This homology appears to include three putative gene families.

## **Introduction**

Telomeres are the DNA-protein complexes at the ends of linear chromosomes. In most eukaryotes, the DNA component consists of tandem copies of a short T/G-rich sequence (typically 5-8 bp) and is maintained by the enzyme telomerase (Wellinger and Sen, 1997; Weilbaecher and Lundblad, 1999). Many proteins are part of the telomere complex and function to regulate telomerase's access to the end and to preserve the protective capping function of telomeres (Zakian, 1995; Shore, 1998; McEachern et al., 2000)

Subtelomeric regions, or telomere-associated sequences, are the regions immediately adjacent to the telomeric repeat arrays (Pryde et al., 1997; Wellinger and Sen, 1997).

Subtelomeric regions are often similar among multiple chromosome ends, and many aspects of structure and function of these regions have been subjects of study in a variety of organisms

including repetitive elements, degenerate telomeric repeats, gene families, gene silencing, and retrotransposons that show affinity for chromosome ends.

The subtelomeric regions of *S. cerevisiae* are the best characterized of any organism to date (Pryde et al., 1997). Two highly repetitive subtelomeric elements are known. Internal to the ~350 bp of telomeric repeats, Y' elements are commonly found at about half of all chromosome ends. These elements are highly conserved, typically 5.4-6.7 kb in size, and are present in 1-4 tandem copies per telomere (Louis, 1995). Internal to the Y' element is the X element, a mosaic element that is more variable than the Y' element. It is composed of a 475-bp consensus core and a variable outer portion composed of families of tandem degenerate repeats (Louis et al., 1994). Internal to the X element, there is often a proximal subtelomeric domain of varying homology among the ends extending for 2-30 kb in strain S288C. A number of different proximal domains are present, each shared between two or three telomeres.

A function that commonly has been attributed to subtelomeric regions is providing a genomic area of greater plasticity. Subtelomeric regions often contain 'contingency genes', genes that have high mutation rates and function in adapting the host to changes in environmental conditions (Barry et al., 2003). In the malaria parasite *P. falciparum*, genes for surface antigens are found in subtelomeric regions. There are 50 to 150 members of this family, most of which are located near the ends of the chromosomes. It has been shown that recombination between non-homologous chromosomes (ectopic recombination) can occur at a high rate during meiosis in *P. falciparum* to create new *var* genes, which would enable efficient evasion of hosts' immune systems (Freitas-Junior et al., 2000). Subtelomeric gene families involved in host immune evasion are also found in a number of other pathogens (reviewed in Barry et al., 2003). Even in non-pathogens such as *S. cerevisiae*, large and variable subtelomeric

gene families are important for growth under particular conditions (Carlson et al., 1985; Naumov et al., 1994; Naumov et al., 1995; Ness and Aigle, 1995; Codon et al., 1998). In humans, genes for olfactory receptors are often subtelomeric and highly variable between different individuals in the chromosome ends in which they reside (Trask et al., 1998; Mefford et al., 2001).

Subtelomeric regions have also been shown to be prone to changes when telomere function is compromised. Recombination in subtelomeric regions can be greatly increased in *K. lactis* cells that lack a functional telomerase (McEachern and Blackburn, 1996). Similarly, deletions that include subtelomeric sequences are common in *S. cerevisiae* when telomerase is absent and telomeres become critically short (Hackett et al., 2001). In humans, mutations in subtelomeric regions are associated with certain diseases including facioscapulohumeral muscular dystrophy (van Overveld et al., 2000) and  $\alpha$ -thalassemia (Flint et al., 1996).

In this study, we have cloned and characterized subtelomeric regions from ten of twelve *K. lactis* telomeres. Our results show that *K. lactis* subtelomeric sequences are unique but share many similarities to those from other organisms.

## **Material and Methods**

### **Strains**

All cloning of subtelomeric sequences was done from *K. lactis* strain 7B520 (Wray et al., 1987). The comparative hybridization study was performed using strains from the Spanish Type Culture Collection (CECT) and other available strains. See Table 2.1 for details.

### **Constructing pKN2**

A 1.5 kb *Bam*HI-*Bgl*III fragment, from pMH3-Tel NoST, containing *S. cerevisiae* *URA3* and 11.5 *K. lactis* telomeric repeats (Natarajan et al., 2003) was cloned into the *Bam*HI site of

low-copy plasmid, pWSK29 (Wang and Kushner, 1991). The resulting plasmid, pKN2 (Figure 2.1), had low-copy number in order to improve the stability of large, repetitive sequences in *E. coli*. The insert was positioned in a way that facilitated cloning of subtelomeric sequence once the plasmid integrated into a *K. lactis* telomere.

#### Yeast transformation

Uncut pKN2 plasmid was transformed into *K. lactis* using a protocol similar to established ones for *S. cerevisiae*. A culture of cells in 1 ml of YPD was grown, and after overnight growth, an additional 1 ml of YPD was added, and the cells were grown for 1 hour more. This culture was pelleted and washed with 500  $\mu$ l cold water. After pelleting and resuspending in 80  $\mu$ l water, 10  $\mu$ l each of 10X TE buffer (pH 7.5) and 1 M lithium acetate (pH 7.5) was added to the cells. This suspension then was incubated at 30° C for 1 hour with and addition of 2.5  $\mu$ l 1 M DTT after 45 minutes. Four hundred  $\mu$ l of cold water was added before pelleting the cells at 4° C. The cells were then washed two times in 250  $\mu$ l cold water and once in 100  $\mu$ l cold 1 M sorbitol with pelleting at 4° C between washes. Cells were suspended in 30  $\mu$ l cold 1 M sorbitol, before adding ~ 1  $\mu$ g pKN2 DNA and electroporated at 1500V (using an Eppendorf electroporator 2510). Five hundred  $\mu$ l 1 M sorbitol was added, and the cell suspension was plated onto synthetic defined (SD) plates lacking uracil and containing 1 M sorbitol.

#### Screening and hybridizations

Two single digests (*Eco*RI and *Pvu*II) of yeast genomic DNA were performed to determine the location of the integrated plasmid. Digests were run on 0.8% agarose gels and

stained with ethidium bromide. Southern blotting was performed using Hybond N<sup>+</sup> membrane (Amersham Pharmacia, Piscataway, NJ). Hybridizations were carried out in Na<sub>2</sub>HPO<sub>4</sub> and SDS (Church and Gilbert, 1984). Washes were carried out in 200 mM Na<sub>2</sub>HPO<sub>4</sub> and 2% SDS for the oligo probe and in 100 mM Na<sub>2</sub>HPO<sub>4</sub> and 2% SDS for all other probes. Membranes were stripped for one hour with 0.4 N NaOH between hybridizations.

The *K. lactis* telomeric oligo Klac 1-25 (ACGGATTTGATTAGGTATGTGG TGT) was used in all telomeric hybridizations. Temperatures of hybridizations and washes were between 50° C and 55° C. The vector-specific probe KN2-220 was a purified 220 bp *Pvu*II-*Sac*I fragment from pBLUESCRIPT SK<sup>®</sup>. This probe was hybridized at 55° C.

The 0.7-kb *Eco*RI-*Xba*I subtelomeric probe was isolated from *K. lactis* telomeric clone KL11B (McEachern and Blackburn, 1994) and hybridized at 55° C. The specific-subtelomeric probe fragments described in Table 2.2 were gel-purified from the rescued low-copy plasmids after digestion with appropriate enzymes then ligated into prepared pBLUESCRIPT SK<sup>®</sup> (digested and treated with (shrimp alkaline phosphatase (Boehringer Mannheim, Indianapolis, IN)) to enable efficient purification for probes. These ligations were transformed into XL-1 supercompetent *E. coli* (Stratagene, La Jolla, CA), and resulting colonies were screened for interruption of the β-galactosidase gene of the vector (blue-white screening) and checked for inserts. Inserts of positive subclones were gel-purified for probes. These were hybridized at 65° C then 50° C or 55° C, if no hybridization was observed at 65° C. All probes except the telomeric probe were prepared using PRIMEIT II labeling kit as recommended by the manufacturer (Stratagene, La Jolla, CA).



## Plasmid Rescue

Genomic DNA from each type of telomeric insert was digested with each restriction enzyme with a site in the polylinker of pKN2 upstream of the *URA3*-tel insert (see Figure 2.1 and Table 2.2). Hybridization data from these digests were used to determine the enzyme that would allow the cloning of the largest amount of subtelomeric sequence. Genomic DNA from each *K. lactis* clone with integrated pKN2 was digested with the enzyme chosen for that clone. After inactivation of the restriction enzyme, approximately 0.5-1 µg DNA from the digest was ligated overnight at 15° C.

A portion of each ligation was transformed into XL-1 supercompetent *E. coli* (Stratagene, La Jolla, CA). Plasmids from transformants were verified by size of insert and by hybridization to the *K. lactis* *EcoRI-XbaI* subtelomeric probe described above.

## Sequencing and analysis

Automated fluorescent dideoxy sequencing was performed by the Molecular Genetics Instrumentation Facility at the University of Georgia. Primers ranging from 18-20 nucleotides were used in primer walking to obtain all sequence reported here.

Sequences were assembled and open reading frames were identified using Sequencher™ 4.1 (Gene Codes Corporation, Ann Arbor, MI). Sequences were subjected to BLASTN and BLASTX analysis (<http://www.ncbi.nlm.nih.gov:80/BLAST/>) (Altschul et al., 1997) against databases to discover homologs or other features. Multiple alignments were performed using the Clustalw tool provided by the European Bioinformatics Institute (EBI- <http://www.ebi.ac.uk>) and adjusted by sight and use of BLAST2 pairwise alignments.

## Results

### Cloning *K. lactis* subtelomeric regions

To clone *K. lactis* subtelomeric sequences, pKN2, a plasmid containing an *S. cerevisiae* *URA3* gene and 287 bp of *K. lactis* telomeric sequence was first constructed. This plasmid contained only 3 bp of *K. lactis* subtelomeric DNA other than the telomere so that it would integrate via recombination within the telomere, enabling cloning of adjacent subtelomeric sequences. This technique was similar to that used previously in cloning subtelomeric regions of *S. cerevisiae* (Louis and Borts, 1995).

Seven independent transformations of pKN2 into *K. lactis* 7B520 were performed in an attempt to recover at least one integration event at telomeres of each of the twelve telomeres present on the six chromosomes of the haploid genome. A total of 41 transformants of pKN2 into 7B520 were recovered and analyzed by Southern blotting. Digestion with *EcoRI* and *PvuII* was used to determine whether transformants contained the expected inserts. The *PvuII* digest was especially helpful since it separated the twelve telomeric restriction fragments into ten different-sized bands, the greatest number seen with any of a large number of different restriction enzymes (data not shown). Transformants were not identified by chromosome of integration, so the names for the chromosome ends used in this study were derived from the *PvuII* pattern, with P1 referring to the smallest *PvuII* telomeric fragment and P12 referring to the largest. The P2 and P3 telomeric fragments migrate together as a doublet, and the P8 and P9 telomeres run together as a second doublet. The *EcoRI* digests confirmed results of the *PvuII* digest and resolved the P2/P3 doublet in the *PvuII* pattern. The P8 and P9 telomeres migrated as a doublet in both digests.

Southern blots of pKN2 transformants were analyzed by hybridizing telomere and vector-specific probes to genomic DNA of transformants that was digested separately with *EcoRI* and with *PvuII*. Our results showed that all of the transformants had the expected restriction patterns for inserts at native telomeres. Figure 2.1 shows Southern blots of *PvuII*-digested DNA hybridized to a telomeric probe (panel B) or to a 220-bp vector-specific probe (referred to as KN2-220; panel C) as well as *EcoRI*-digested DNA hybridized to the telomeric probe (panel D). The position within the plasmid vector of the sequence used as the KN2-220 probe is indicated by the thick black box in Figure 2.1A. This region of the vector is the only vector-derived sequence that remained attached to subtelomeric sequences after digestion with *PvuII* and therefore detected the particular subtelomeric DNA fragment from a transformant that had integrated pKN2 (see PS and ES fragments in Figure 2.1A). A shift of one band in the telomeric restriction pattern results from the integration of the plasmid into a telomere. *PvuII* digestion was expected to cut off most of the vector sequence leaving about 0.2 kb connected to the chromosome end whereas an *EcoRI* digest leaves 5 kb of vector sequence attached to the subtelomeric fragment. This accounts for the different shifts seen in the two digests. These shifted bands are also often expected to be sharper than true telomeric bands since they are internal fragments as opposed to telomeric ends that are heterogeneous in length due to differences in length of telomeric DNA. This was clearly seen for smaller fragments such as the P2 and P3 inserts (Figure 2.1B). Another novel fragment expected in transformants is the terminal fragment (PT and ET in *PvuII* and *EcoRI* digests, respectively; Figure 2.1A) that contains the telomere and *URA3*. This is a 1.5 kb band in both the *PvuII* digest and the *EcoRI* digest. In the *EcoRI* digest (Figure 2.2D) it is obscured by other telomeric fragments, but it was seen clearly with a *URA3* probe (data not shown).

Hybridization with telomeric and KN2-220 probes also revealed the presence of tandem copies of pKN2 sequence at telomeres in most of the transformants. In a *PvuII* digest, extra copies of pKN2 were detected as a band at 2 kb (PI band marked with arrow in Figure 2.1B, C) and in *EcoRI* digests as a band at ~6 kb (EI band in Figure 2.1D). This second band was present in 38 of the 41 transformants. Figure 2.1A shows the expected structure of a dimeric integration of pKN2. We have not determined whether the tandem inserts represent dimers or higher order multimers. However, the varied relative intensity of PI and PS bands between different transformants (Figure 2.1C) suggests that multimeric inserts were likely present.

Characterization of the transformants showed that 25 of them had pKN2 inserted in one of the six telomeres found in the smallest group of *EcoRI* fragments (P2, P4, P7, P10, P11, or P12). Inserts into other telomeres included one at P6, nine at P3, two at P5, and two at P1. None of these transformants showed plasmid integration at the telomeric band P12. Only two were at the P8-9 doublet (labeled as P9 throughout this work). Sequence data from the telomeric end to the *EcoRI* site as well as restriction data using 12 restriction enzymes suggested that these two transformants were either the same end or two indistinguishable ends.

To clone subtelomeric sequence from the transformants, ten restriction enzymes were considered for use. Each of these enzymes cuts once within the polylinker of an integrated pKN2 and permits excision of subtelomeric sequence still attached to the plasmid vector (Figure 2.1A). Such fragments could then be circularized by ligation and transformed into *E. coli*. The choice of enzyme(s) to use for cloning of each end was based on Southern data (data not shown); transformants were digested by all ten enzymes and hybridized to the KN2-220 probe. For each digest, the anticipated size of the extra band that resulted from the additional copy of the vector was calculated. The sizes differed for each digest due to restriction sites within *URA3*. This left

the other band resulting from each digest to be used to estimate the amount of sequence that would be cloned using that digest.

The tandem integration of pKN2 in most transformants was of special concern in the plasmid rescue procedure. During plasmid rescue many circles may be created, but only those derived from the vector plasmid pKN2 would enable *E. coli* to grow in presence of ampicillin. However, with two or more copies of vector sequence at a given end, a functional plasmid may be either one that contains *K. lactis* subtelomeric sequence (e.g., the ES fragment shown in Figure 2.1A) or simply a recircularized copy of the original pKN2 plasmid (e.g., the EI fragment shown in Figure 2.1A), both of which would yield white colonies in a blue-white  $\beta$ -galactosidase screen. Although producing an undesirable background of unwanted *E. coli* transformants, this situation was of some use since the tandem copy could serve as a positive internal control for the plasmid rescue conditions. Recovery of pKN2 without successful recovery of the expected plasmid with a particular piece of subtelomeric DNA from a *K. lactis* transformant could be taken as indicating that the latter plasmid might not be recoverable.

A number of plasmids from each end (P6, 6 plasmids, fewest – P1, 70 plasmids, most) was examined before recovering one from each transformant that appeared correct based on size of insert and hybridization to a subtelomeric probe that is known to hybridize to eleven of the twelve telomeres in 7B520. *EcoRI* was used successfully as a recovery enzyme for all ends tested. The P1 clone contained an extra *EcoRI* fragment that later proved to not be subtelomeric in origin. *EcoRI* telomere fragments ranged 0.7-3 kb, which likely contributed to the consistent recovery of plasmids that contained subtelomeric sequence. An average of 35% of plasmids recovered after cleavage of genomic DNA with *EcoRI* were found to contain the expected subtelomeric DNA (with the remainder being recircularized pKN2).

For nine of the ten telomeres for which we had obtained pKN2 integrations, we were also able to recover substantially larger subtelomeric pieces using other restriction enzymes. Six telomeres have been recovered as fragments ranging 9-18 kb, and the other four had 0.7-4 kb of subtelomeric sequence cloned. For P2, a plasmid larger than the *EcoRI* plasmid was not recovered even after screening 193 plasmids using the enzymes *BamHI*, *ClaI*, and *XmaI*. Our results with subtelomeric cloning are summarized in Appendix A. In all cases described in this work, recovered subtelomeric sequences were found to match the sizes of the corresponding native terminal restriction fragments. This indicates that gross rearrangements in the cloned sequences are unlikely.

#### Structure of *K. lactis* subtelomeric regions

The four 7B520 telomeres from which we recovered 0.7-4 kb of subtelomeric sequence were completely sequenced. The six telomeres recovered with much larger regions of subtelomeric DNA were in most cases just partially sequenced. However, at least 1.2 kb of the sequence immediately adjacent to the telomeric repeats was sequenced from each. One of these large telomeric clones, P10, was sequenced over its entire 9.8 kb length. The regions sequenced in our subtelomeric clones are shown in Figure 2.2.

Previous hybridization experiments have shown that all but one telomere from *K. lactis* 7B520 (the P1 telomere) share close sequence similarity in the region ~ 0.6 kb immediately adjacent to the telomeric repeats (McEachern and Blackburn, 1996). Our results here confirm and extend those results. The homology shared between subtelomeric regions of at least two telomeres is shown in Figure 2.2 as colored boxes. At least 8 of the 12 telomeres share >85% homology extending at least 1.5 kb from the telomeric repeats. This number may be as high as

11 of 12 as the P8 and P12 telomeres were not cloned, and the P2 telomere was recovered only as a fragment with ~0.7 kb of subtelomeric DNA.

A number of the *K. lactis* telomeres share subtelomeric homology that extends multiple kilobases from the telomeres. Of five telomeres with at least some sequence information from positions 8-18 kb internal from the telomeric repeats, three have highly homologous sequence that is shared with at least one of the others and extends to the internal end of the sequenced region. The telomere clone with the most available sequence, P10, contains little or no unique DNA sequence in its 9.8 kb length.

Five of the ten cloned telomeres, P1, P3, P4, P5, and P11, had appreciable lengths of sequence that are not shared with other cloned telomeres of this study. These sequences are shown as plain black lines in Figure 2.2. In each case, this sequence extends to the internal border of the sequenced region and represents potentially unique sequence. The P1 subtelomeric region, is unique in that it lacks strong homology to subtelomeric regions from any of the other telomeres.

Subtelomeric sequences located within 1.5-2 kb from the base of the telomeric repeat arrays were distinguishable from more internal subtelomeric sequence in a number of respects. As noted above, they are homologous among a greater number of chromosome ends than are more internal subtelomeric regions. They also show little or no sign of containing genes as ORFs tend to be short and have no significant homology to regions from other organisms. An especially notable characteristic of these immediate subtelomeric regions is a very pronounced strand bias in the content of pyrimidines and purines. As shown in Figure 2.3, the strand running 5' to 3' toward the telomere averages 70% purine over a 1.6 kb region next to the telomeric repeats. The strand bias is nearly 80% purine in the region within 400 bp from the telomere and

gradually declines at more internal positions. The unique P1 telomere exhibits a very similar pattern of strand bias to other telomeres despite having very limited homology to them (Figure 2.3). Among the four types of subtelomeric regions we identified that were non-homologous to each other beyond ~1.6-2 kb from the telomere (represented by the P1, P3, P5, and P10 telomeres), regions more internal than this invariably lack appreciable strand bias in purine content (Figure 2.3).

The sequences of the immediate subtelomeric sequence for each of the ten telomeres we have cloned are shown aligned with one another in Figure 2.4. The sequences are aligned with the telomeric repeats at the start (top left of Figure 2.4). Numbering of the subtelomeric sequences begins with position “1” as the first base pair that does not match the sequence of the telomeric repeat. The sequences shown in Figure 2.4 are the pyrimidine-rich strands. Sequence features present in this region include three families of short, irregular tandem repeats. A series of 8-9-bp repeats is centered ~70 bp from the telomere, and a second series of similar repeats is present ~180 bp from the telomere. Several copies of an unrelated 14-bp direct repeat are present near position 580. Each of these repeat families showed the same pronounced purine/pyrimidine strand bias that characterize the region in general. Most of the larger gaps that occur in the alignment of the subtelomeric regions (excluding P1) are due to variations in the numbers of these repeats. The P1 telomere lacks the short tandem repeats present at other telomeres. Interestingly, although the P1 subtelomere is highly dissimilar from all other cloned subtelomeric regions, it does appear to have weak homology in places to them. The region in P1 from position 280 to 704 (Figure 2.4) has 59% identity to the corresponding regions of other subtelomeres. The DNA in P1 located between this region and the telomere also has small patches of identity. Sequence that is ~0.4 kb internal to the end of the homology between P1



and the other subtelomeres was hybridized to 7B520 genomic DNA. The resulting Southern blot of Figure 2.5A supports the claim that the P1 sequence reported in this study is unique within the 7B520 genome. A single fragment hybridized to the probe in each of eight different restriction digests that were examined.

We have named the ~1.5 kb subtelomeric region adjacent to most of the *K. lactis* telomeres the R element due to its purine-rich nature on the telomeric G-strand. This element is present in conserved form in 11 of 12 telomeres of 7B520. Hybridization using ~600 bp of the R element as a probe to 7B520 DNA cut with multiple restriction enzymes is seen in Figure 2.5B. The P1 telomere can be said to have a smaller, highly degenerate version of the R element.

In addition to recovering subtelomeric sequences from different telomeres, our cloning experiments also recovered the adjacent telomeric repeats. Because of our cloning strategy, it was expected that some of the recovered telomeric repeats would be derived from pKN2. Those nearest to the subtelomeric sequence, however, are likely to be the native repeats as they were present in the cell prior to pKN2 integration. Among 92 native telomeric repeats cloned from ten subtelomeres, all but one had the wild type 25 bp sequence. The single variant repeat observed was a 26-bp repeat resulting from a single-bp insertion within three consecutive T residues (TTTTGATTAGGTATGTGGT GTACGGA with extra base underlined). This type of rare variant repeat has been noted before in *K. lactis* (Tzfati et al., 2000).

Subtelomeric sequence from four of the cloned ends shows sequence match to the *K. lactis* retrotransposon Tk11.1, a TyI-like element (Neuveglise et al., 2002). Figure 2.6 shows a schematic of the alignment of Tk11.1 with the sequences present in the different subtelomeric regions. P3 has the greatest amount of similarity: an ~2 kb stretch of 96% identity with the pol region of the element. On either end of this alignment, the homology ends, indicating that the P3

Tkl1.1 sequence is an incomplete element. A probe made from the P3 subtelomere (described in Appendix A and mapped in Figure 2.2), located completely within this pol homology is shown hybridized to 7B520 DNA cut with each of multiple restriction enzymes in Figure 2.5C. Most of these enzymes produced only one band that hybridized to the *Pst*I Tkl1.1 fragment probe. As the nearest sites for these enzymes reside outside of Tkl1.1 sequence on one or both sides of the element, inserts at positions other than the specific site mapped in the P3 telomere were expected to produce additional bands on the gel. The single enzyme (*Nsi*I) that produced two bands has a site within the fragment used as a probe. Our results therefore indicate that the full length of the sequence used as probe is present in the 7B520 genome only at the P3 subtelomeric location we have identified.

The other telomeres with Tkl1.1 sequence only have LTR sequences. P10 and P9 have an approximately 400-bp stretch of 86% identity to the 3' LTR of the element, P4 has a 142-bp stretch at 95% identity to the 5' LTR, and P5 has a 182-bp stretch of 78% identity to the 3' LTR. The two LTRs of the originally reported Tkl1.1 element are only 93% identical to each other and can therefore be distinguished from one another. A hybridization probe made from an ~1kb *Eco*RI fragment from P7 (Figure 2.2 and Appendix A), which includes the Tkl1.1 LTR, produced multiple bands when hybridized to 7B520 DNA (Figure 2.5B). These bands represent at least three subtelomeric sequences and indicate that either the LTR or the other subtelomeric sequence present in the probe is present at a number of other sites in the genome.

#### ORFs found in *K. lactis* subtelomeric regions

There are multiple regions in the *K. lactis* subtelomeric regions we have sequenced that have homology to genes from other organisms. Figure 2.2 shows positions of ORFs of >100

amino acids and of regions with homology to ORFs from *S. cerevisiae* or from other organisms. An ~2 kb region on P1 has amino acid homology with three related genes, *FLO1*, *FLO5*, and *FLO9*, a subtelomeric family of *S. cerevisiae* genes involved in cell adhesion known as flocculation. This similarity appears to be chiefly to the region containing copies of flocculin repeat A, a 45-amino acid repeat. There are one, eight, and eighteen copies of this repeat in the three *S. cerevisiae* genes (Teunissen and Steensma, 1995), and at least nine are found in the *K. lactis* *FLO* gene. Because our sequence ends within the flocculin repeat region, the total number of repeats is not known. Of the nine we have sequenced, six complete repeats ranging from 44-47 amino acids in length, and three are partial repeats of 16, 24 and 27 amino acids. The *K. lactis* flocculin repeats are generally about 50% identical to their *S. cerevisiae* counterparts. Downstream of the flocculin A repeats in the *S. cerevisiae* *FLO1*, *5*, and *9* genes, there are several serine-rich repeats (Teunissen and Steensma, 1995). In this region of the *K. lactis* gene, there is also serine-rich sequence similarity that continues through the end of the alignment.

There are regions on P10 and P7 nearly 10 kb from the telomere that share amino acid homology with the putative arsenite transport protein Arr3p. These P10 and P7 sequences are 99% identical with each other on the nucleotide level. On the protein level there is 67% identity and 81% similarity between the *S. cerevisiae* and *K. lactis* copies and 63% identity at the nucleotide level. The available *K. lactis* sequence ends at a point in the alignment 20 amino acids from the 3' end of the *S. cerevisiae* gene.

There are four *K. lactis* subtelomeric sequences that appear to encode proteins with homology to *S. cerevisiae* Mch2p, a protein of unknown function (Figure 2.2). These homologies are present on P6, P7, P9, and P10. Three of the four are in regions of incomplete sequence. Pairwise comparisons between each translated sequence and the *S. cerevisiae* Mch2p

show ~45% homology. The four *K. lactis* sequences show 98-99% identity with each other at the nucleotide level.

BLASTX analysis of P11 internal sequence showed a 55-amino acid alignment (50% identity/ 69% similarity) containing a 27-amino acid protein kinase C-terminal domain (InterPro Accession Number IPR000961E). This domain is present in twelve yeast cAMP-dependent kinases such as Tpk1p and in kinases of other organisms including humans. The homology in P11 is close to an end of available sequence.

BLASTX analysis also showed that five of the *K. lactis* subtelomeric sequences have regions with homology to hypothetical oxidoreductase proteins from several organisms. These hypothetical proteins are believed to be oxidoreductases based on their homology to members of a dioxygenase superfamily (Aravind and Koonin, 2001). The highest score was assigned to the alignment with hypothetical protein 15E6.100 from *Neurospora crassa*. Two of the five *K. lactis* ORFs (in the P4 and P7 sequences) have internal 250-bp deletions (shown with double asterisks in Figure 2.2).

There are four other open reading frames throughout the recovered sequence that would yield peptides of at least 100 amino acids (Figure 2.2). None of these have significant alignments to any known sequences when analyzed with BLASTX. The ORF within the R element in P7 is not present as an ORF of >100 amino acids in the other homologous regions due to base pair changes that generate stop codons.

Probing other *Kluyveromyces* strains for homology to subtelomeric sequences from 7B520

A selection of strains, many from the Spanish Type Culture Collection (CECT), has been used to assess whether subtelomeric sequences from *K. lactis* 7B520 identified in this study are

present in other *Kluyveromyces* strains and species. The positions of the DNA fragment probes made from 7B520 DNA are indicated in Figure 2.2. The P1 probe, 1012 bp located 1.2-2.2 kb from the telomere, hybridized to all but one of the *K. lactis* strains, to one of the *K. marxianus* strains, and to the *K. wickerhamii* strain (Figure 2.7A). *EcoRI*-digested DNA from five of the six *K. lactis* strains showed hybridization to an ~3.5 kb band identical in size to the P1 *EcoRI* fragment of 7B520. The remaining *K. lactis* strain showed no hybridization to this probe. Additionally, one of two *K. marxianus* strains showed hybridization to the same ~3.5 kb band as well as to two additional bands above 10 kb. The hybridization of the P1 probe to genomic DNA from at least some *K. marxianus* strains is consistent with Génolevures BLAST data; the P1 sequence shows homology to three telomeric *K. marxianus* RSTs (data not shown). The 932-bp *PstI* fragment probe from P3, located about 3 kb from the telomere, hybridized to a single band in all *K. lactis* strains and to three bands in the *K. dobzhanskii* strain (Figure 2.7B).

A telomere-adjacent probe (R1-6 *EcoRI*-*XbaI*) from the R element was also used in examining these strains (Figure 2.7C). This hybridization showed a series of diffuse bands recognizable as telomeric bands in each *K. lactis* strain examined except *K. lactis var. drosophilarum* 10390 which showed a series of faint and generally larger bands. *K. dobzhanskii* showed several bands that hybridized well to the probe that were different in size than the *K. lactis* bands. The *K. wickerhamii* strain and both *K. marxianus* strains showed no hybridizing bands. Two DNA fragments from P10 (Appendix A) were also used as hybridization probes to the various strains. These hybridized to the same or very similar pattern of at least five bands ranging from 1 kb to about 8 kb in 7B520 and in all *K. lactis* strains except *K. lactis var. drosophilarum* 10390 which showed no hybridization (data not shown).

The overall picture from the hybridizations of 7B520 subtelomeric probes to other strains is consistent with the 5.8S rRNA phylogeny study previously performed using 39 CECT strains including the eight used in this study (Belloch et al., 2002). *K. dobzhanskii* appears to be the most closely related to *K. lactis* in both studies and *K. wickerhamii* the most distantly related.

## Discussion

The *K. lactis* subtelomeric regions presented in this study contain several features that are shared with subtelomeric regions of other organisms. One such characteristic is the presence of two general zones of homology that are shared between multiple ends (Pryde et al., 1997). Distal domains are generally shorter, can lack functional genes, and are present at a large percentage of telomeres. Proximal domains of homology are usually longer, contain genes, and are present at fewer ends. Multiple unrelated proximal domains can be present within a given organism.

An example of two zones of homology in humans is seen in hybridization analysis using 4q subtelomeric sequence. A 15-kb region just proximal to the 4q telomere is shared with 17 other ends, and sequence internal to this region (15-60 kb from the end) was only shared with four other ends (Flint et al., 1997). The distal subdomain contains a high density of ESTs and short sequences that match other ends in a patchwork fashion (Der-Sarkissian et al., 2002).

In *S. cerevisiae* a related picture is seen on a smaller scale. The distal domain of homology includes the X and Y' elements. The X element is present in some form at all *S. cerevisiae* ends and consists of a core sequence of 472 bp and up to several hundred base pairs of families of variable repeats (Louis et al., 1994). The Y' element, often present at about half of telomeres, consists of a 5-6 kb sequence that is sometimes present in more than one copy per

telomere (Pryde and Louis, 1997). The larger blocks of homology that make up the proximal domains of *S. cerevisiae* subtelomeres exist in ten distinct sequence groups with common blocks of genes shared between two to three telomeres (<http://www.le.ac.uk/genetics/ejl12/research/telostruc/ClustersSmall.html>, Ed Louis, personal communication). Twenty-three of 32 telomeric ends are part of these proximal domain groups in *S. cerevisiae* S288C.

*K. lactis* 7B520 can also be described as having two zones of homology among its telomeric ends. The R element, which lacks detectable homology to the X and Y' elements of *S. cerevisiae*, comprises the distal domain and is present on eleven of twelve telomeres in conserved form and on the twelfth in degenerate form. In this study, proximal domains of homology, consisting minimally of a putative oxidoreductase gene, have been observed to be present on at least five telomeres. The limited sequencing done in this study prevents us from determining the full extent of proximal domains of homology at telomeres in 7B520. However, it does appear likely that *K. lactis* will differ from *S. cerevisiae* in some respects. *K. lactis* is unlikely to have as many different types of proximal homology domains as *S. cerevisiae*. Thus far, not more than a single type of proximal domain can be said with certainty to exist in the 7B520 strain. *K. lactis* also has a higher percentage of telomeres with the same proximal domain (at least 5 of 12 telomeres) than is seen in the sequenced *S. cerevisiae* strain.

Another feature in the subtelomeric sequences of many organisms is the presence of smaller repetitive elements. A variety of such repeats has been observed in different organisms. A well-studied example in *S. cerevisiae* is the family of subtelomeric repeats (STRs) found in the distal part of the X element. Four types of STR elements have been characterized. They vary in length (35-150 bp), but all are made up of short degenerate repeats. The *K. lactis* R element lacks

homology to the STRs of *S. cerevisiae*, but it does have its own small repeats in the distal ~650 bp of the R element. These are similar in size (~6-14 bp) to those within *S. cerevisiae* STRs and occupy a similar relative position in the distal part of their respective element. Human subtelomeric regions have been observed to be an area where minisatellite sequences are clustered (Royle et al., 1988; Wells et al., 1989; Vergnaud et al., 1991). Due to the observations of major clusters at several chromosome ends and of the suspected transposition of sequence containing a minisatellite from a terminal region to an internal region (Wong et al., 1990), it has been suggested that this region even serves as an origin for these minisatellites, which are then spread to other areas of the genome (Amarger et al., 1998).

Telomere-like repeats have also been observed in subtelomeric DNA. In human cells, the human telomeric repeat, TTAGGG and degenerate copies of it, are often located at the border between the two zones of homology described earlier (Flint et al., 1997). *S. cerevisiae* subtelomeres can contain two types of telomere-like sequences: tracts of sequence matching its own variable telomeric repeats and short tracts of the vertebrate telomeric repeat (TTAGGG). TTAGGG sequences, both single and multiple copy, have been shown to be binding sites, *in vivo* (Koering et al., 2000) and *in vitro* (Brigati et al., 1993), for the essential yeast protein Tbf1p and can function as a telomere in *S. cerevisiae* (Henning et al., 1998; Alexander and Zakian, 2003). In *K. lactis*, no tandem blocks of TTAGGG repeats were found in any of the sequenced subtelomeric DNA (data not shown). Single TTAGGG sequences were found at the same location (~900 bp from the telomere) in eight of the ten ends reported. At one telomere (P1) a TTAGGG sequence was at a different position and in the opposite orientation of the others, and in the cloned P2 subtelomere, the sequence was not present.



Strand bias in base composition is another characteristic of *S. cerevisiae* subtelomeric sequences that we have observed in *K. lactis*. In *S. cerevisiae*, on the strand reading 5' to 3' toward the telomere, X elements have a high G/C ratio and a somewhat low A/T ratio and Y' elements have relatively high G/C and A/T ratios. It has been proposed that this strand asymmetry in base composition is a result of replication-associated mutation (Gierlik et al., 2000). DNA replication has been implicated in the bias that is observed in the base composition in many bacteria as all bacterial genomes with well-defined replication origins and termini are asymmetric in the base composition of leading and lagging strands (Lobry, 1996; Grigoriev, 1998). Strand bias in replicative mutagenesis has recently been shown associated with yeast origins as well (Pavlov et al., 2002). Eukaryotic telomeres may be similar to bacterial chromosomes since in both cases leading and lagging replication strands are set due to their positions to the ARS. In support of this, the strand bias that exists in *S. cerevisiae* subtelomeric sequences is diminished internal to positions of ARS elements (Gierlik et al., 2000). This might predict that the position marking the beginning of strand bias in base composition of the R element and the degenerate P1 subtelomeric sequence of *K. lactis* may indicate the position of an origin of replication. Alternatively, or in addition, the strand bias in base composition of *K. lactis* subtelomeric sequences might be due to selection for some advantageous property, one possibility of which is discussed below.

Another feature of *K. lactis* subtelomeric sequences that has been seen in other organisms is the presence of transposable elements (Zou et al., 1995; Flint et al., 1996; Mefford et al., 2001). An extreme example of the association between telomeric ends and transposons is the *D. melanogaster* chromosome end, which is composed of, and maintained by, active retrotransposons instead of DNA synthesis by telomerase. *S. cerevisiae* has a retrotransposon

associated with its ends; Ty5 is an element that is typically found near the X element. Active copies of this element have been shown to have a preference for insertion at this region of the genome (Zou et al., 1996). Fragments of a characterized *K. lactis* LTR-retrotransposon (Tk11.1) are found near at least four telomeres of 7B520. Tk11.1 is an incomplete Ty1-like element that has a truncated gag element (Neueglise et al., 2002). LTR fragments from this element and from a truncated version (Tk11.2) are believed to be present in the genome in at least 30 copies based on random sequence tags from the Génolevures project (Neueglise et al., 2002). Our comparisons of Génolevures BLAST data of a *K. lactis* telomere search and a Tk11.1 LTR search showed that at least four of the 30+ copies in strain CLIB 210 are subtelomerically located (data not shown). Three of these sequences appear to correspond to inserts reported in this study.

Subtelomeric regions have been described as the most plastic of the genome (Pryde et al., 1997). Manifestations of this include a lack of sequence conservation between species, allelic variability within species, and the flexible nature of many of the resident genes. Genes located in subtelomeric regions often fit the definition of contingency genes, genes that have a high mutation rates and can help the host organism adapt to a changed environment (Moxon et al., 1994; Barry et al., 2003).

The subtelomeric surface antigen genes of *P. falciparum*, *T. brucei* (Borst and Ulbert, 2001) and other parasites are well known examples of eukaryotic contingency genes (Barry et al., 2003). In human cells, variation is prominent among the subtelomeric members of the olfactory receptor (OR) gene family. In light of these results and known properties of subtelomeric regions, this region of the human genome has been put forth as a “nursery” for new olfactory receptor proteins (Mefford et al., 2001).

Some gene families near telomeres in *S. cerevisiae* are clear candidates for being contingency genes (Barry et al., 2003). In industries such as brewing, baking, and wine-making, specific properties such as ability to ferment particular sugars, tolerance to molasses, and flocculation, have undergone strong selection for centuries (Codon et al., 1998; Dequin, 2001). Substantial variation in subtelomeric genes controlling some of these traits has been observed. Many industrial *S. cerevisiae* strains encode subtelomeric *SUC* and *RTM* genes, which are involved in the ability to utilize sucrose and resistance to molasses, respectively (Carlson et al., 1985; Ness and Aigle, 1995). These genes vary in copy number and telomere of location between strains, and they are completely lacking from some lab and industry strains (Naumov et al., 1996; Denayrolles et al., 1997). Similarly, the subtelomeric *MAL* and *MEL* gene families are involved in utilization of other sugars, and they too, vary in number between strains (Naumov et al., 1994; Naumov et al., 1995).

*S. cerevisiae* and *K. lactis* have similarities in genes that are found in subtelomeric DNA as *FLO*, *ARR3* and *MCH2* family genes are present there in both yeasts. The *FLO* and *ARR3* genes have characteristics suggesting they could be contingency genes. Both genes are expected to participate in processes, flocculation and arsenite resistance, respectively, that would affect the cell's response to its environment. Flocculation in *S. cerevisiae* is known to involve *FLO* gene family members and appears to be a variable and often unstable trait (Dequin, 2001; Verstrepen et al., 2003). The function of *MCH2* is unclear. It is homologous to monocarboxylic acid transporters but does not appear to play that role in *S. cerevisiae* (Makuc et al., 2001).

Subtelomeric gene families likely arise as a result of recombination between chromosome ends. New forms of the *P. falciparum* subtelomeric *var* gene copies have been observed resulting from increased ectopic recombination (Freitas-Junior et al., 2000). In *S. cerevisiae*, a

constitutive mutant form of the subtelomeric *MAL63* gene was derived via gene conversion (Wang and Needleman, 1996). Also, subtelomeric recombination clearly occurs between Y' elements in wild type strains (Louis and Haber, 1990). This likely contributes to the homogenization of sequence within these elements. In *K. lactis*, homogenization of subtelomeric sequence also has occurred, given the extensive homology that exists between the ends. The P1 telomere of 7B520, with its poor sequence match to the other ends, may be an example of a sequence that has escaped the homogenization and diverged over time to become a quite distinct sequence.

Recombinational telomere maintenance provides an extreme example of how recombination can rapidly alter regions in and around telomeres. This process has been observed in the absence of telomerase activity in human cell lines (Bryan et al., 1997; Dunham et al., 2000), *S. cerevisiae* (Lundblad and Blackburn, 1993; Teng and Zakian, 1999), and *K. lactis* (McEachern and Blackburn, 1996). The mechanisms involved in maintaining telomeres through recombination have not been fully resolved, but they are believed to include homologous recombination as a means of spreading telomeric sequences to other ends in the cell. In *S. cerevisiae*, Type 1 survivors show greatly amplified arrays of subtelomeric Y' elements. In Type 2 survivors of *S. cerevisiae* and all *K. lactis* survivors examined, the amplification seen is that of telomeric sequence (McEachern and Blackburn, 1996; Teng and Zakian, 1999). In *K. lactis*, the multiple lengthened telomeres within a given survivor appear to arise from a single source, suggesting widespread sequence homogenization (Natarajan, 2002). *K. lactis* survivors also often showed spreading of a subtelomeric marker gene from one telomere to most or all other telomeres in the cell (McEachern and Iyer, 2001).

As mentioned above, the extent of common sequence found at multiple ends of *K. lactis* 7B520 implies that there has been homogenization of these ends over time. The question arises of how a cell determines which chromosome end to use as a donor in subtelomeric gene conversion events. It seems likely that certain sequence features would perform better mechanistically as a recombination donor than others. *In vitro* assays with RecA, a prokaryotic DNA strand exchange protein, and its yeast equivalent, Rad51p, show a preference for both binding GT-rich DNA and promoting strand invasion with GT-rich DNA (Dixon and Kowalczykowski, 1991; Tracy et al., 1997). It is possible that sequence features of the *K. lactis* R element, such as its strand bias in base composition and its families of short repeats, might favor its ability to spread to other telomeres through recombinational processes.

The availability of *K. lactis* subtelomeric sequence will be useful in many future studies. Interesting questions that could be examined include the extent of subtelomeric variation in natural populations and mutants affecting telomere function, the evolutionary roles of subtelomeric genes, and possible functions of the R element.

## References

- Alexander, M. K. and Zakian, V. A. (2003). Rap1p telomere association is not required for mitotic stability of a C(3)TA(2) telomere in yeast. *EMBO J* **22**: 1688-96.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-402.
- Amarger, V., Gauguier, D., Yerle, M., Apiou, F., Pinton, P., Giraudeau, F., Monfouilloux, S., Lathrop, M., Dutrillaux, B., Buard, J. and Vergnaud, G. (1998). Analysis of distribution

- in the human, pig, and rat genomes points toward a general subtelomeric origin of minisatellite structures. *Genomics* **52**: 62-71.
- Aravind, L. and Koonin, E. V. (2001). The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol* **2**: RESEARCH0007.
- Barry, J. D., Ginger, M. L., Burton, P. and McCulloch, R. (2003). Why are parasite contingency genes often associated with telomeres? *Int J Parasitol* **33**: 29-45.
- Belloch, C., Fernandez-Espinar, T., Querol, A., Dolores Garcia, M. and Barrio, E. (2002). An analysis of inter- and intraspecific genetic variabilities in the *Kluyveromyces marxianus* group of yeast species for the reconsideration of the *K. lactis* taxon. *Yeast* **19**: 257-68.
- Borst, P. and Ulbert, S. (2001). Control of VSG gene expression sites. *Mol Biochem Parasitol* **114**: 17-27.
- Brigati, C., Kurtz, S., Balderes, D., Vidali, G. and Shore, D. (1993). An essential yeast gene encoding a TTAGGG repeat-binding protein. *Mol Cell Biol* **13**: 1306-14.
- Bryan, T. M., Englezou, A., Dalla-Pozza, L., Dunham, M. A. and Reddel, R. R. (1997). Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat Med* **3**: 1271-4.
- Carlson, M., Celenza, J. L. and Eng, F. J. (1985). Evolution of the dispersed SUC gene family of *Saccharomyces* by rearrangements of chromosome telomeres. *Mol Cell Biol* **5**: 2894-902.
- Church, G. M. and Gilbert, W. (1984). Genomic sequencing. *Proc Natl Acad Sci U S A* **81**: 1991-5.

- Codon, A. C., Benitez, T. and Korhola, M. (1998). Chromosomal polymorphism and adaptation to specific industrial environments of *Saccharomyces* strains. *Appl Microbiol Biotechnol* **49**: 154-63.
- Denayrolles, M., de Villechenon, E. P., Lonvaud-Funel, A. and Aigle, M. (1997). Incidence of SUC-RTM telomeric repeated genes in brewing and wild wine strains of *Saccharomyces*. *Curr Genet* **31**: 457-61.
- Dequin, S. (2001). The potential of genetic engineering for improving brewing, wine-making and baking yeasts. *Appl Microbiol Biotechnol* **56**: 577-88.
- Der-Sarkissian, H., Vergnaud, G., Borde, Y. M., Thomas, G. and Londono-Vallejo, J. A. (2002). Segmental polymorphisms in the proterminal regions of a subset of human chromosomes. *Genome Res* **12**: 1673-8.
- Dixon, D. A. and Kowalczykowski, S. C. (1991). Homologous pairing in vitro stimulated by the recombination hotspot, Chi. *Cell* **66**: 361-71.
- Dunham, M. A., Neumann, A. A., Fasching, C. L. and Reddel, R. R. (2000). Telomere maintenance by recombination in human cells. *Nat Genet* **26**: 447-50.
- Flint, J., Bates, G. P., Clark, K., Dorman, A., Willingham, D., Roe, B. A., Micklem, G., Higgs, D. R. and Louis, E. J. (1997). Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Hum Mol Genet* **6**: 1305-13.
- Flint, J., Rochette, J., Craddock, C. F., Dode, C., Vignes, B., Horsley, S. W., Kearney, L., Buckle, V. J., Ayyub, H. and Higgs, D. R. (1996). Chromosomal stabilisation by a subtelomeric rearrangement involving two closely related Alu elements. *Hum Mol Genet* **5**: 1163-9.

- Freitas-Junior, L. H., Bottius, E., Pirrit, L. A., Deitsch, K. W., Scheidig, C., Guinet, F., Nehrass, U., Wellems, T. E. and Scherf, A. (2000). Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* **407**: 1018-22.
- Gierlik, A., Kowalczyk, M., Mackiewicz, P., Dudek, M. R. and Cebrat, S. (2000). Is there replication-associated mutational pressure in the *Saccharomyces cerevisiae* genome? *J Theor Biol* **202**: 305-14.
- Grigoriev, A. (1998). Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Res* **26**: 2286-90.
- Hackett, J. A., Feldser, D. M. and Greider, C. W. (2001). Telomere dysfunction increases mutation rate and genomic instability. *Cell* **106**: 275-86.
- Henning, K. A., Moskowitz, N., Ashlock, M. A. and Liu, P. P. (1998). Humanizing the yeast telomerase template. *Proc Natl Acad Sci U S A* **95**: 5667-71.
- Koering, C. E., Fourel, G., Binet-Brasselet, E., Laroche, T., Klein, F. and Gilson, E. (2000). Identification of high affinity Tbf1p-binding sites within the budding yeast genome. *Nucleic Acids Res* **28**: 2519-26.
- Lobry, J. R. (1996). Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol* **13**: 660-5.
- Louis, E. J. (1995). The chromosome ends of *Saccharomyces cerevisiae*. *Yeast* **11**: 1553-73.
- Louis, E. J. and Borts, R. H. (1995). A complete set of marked telomeres in *Saccharomyces cerevisiae* for physical mapping and cloning [published erratum appears in *Genetics* 1995 May;140(1):420]. *Genetics* **139**: 125-36.
- Louis, E. J. and Haber, J. E. (1990). Mitotic recombination among subtelomeric Y' repeats in *Saccharomyces cerevisiae*. *Genetics* **124**: 547-59.



- Louis, E. J., Naumova, E. S., Lee, A., Naumov, G. and Haber, J. E. (1994). The chromosome end in yeast: its mosaic nature and influence on recombinational dynamics. *Genetics* **136**: 789-802.
- Lundblad, V. and Blackburn, E. H. (1993). An alternative pathway for yeast telomere maintenance rescues *est1*- senescence. *Cell* **73**: 347-60.
- Makuc, J., Paiva, S., Schauen, M., Kramer, R., Andre, B., Casal, M., Leao, C. and Boles, E. (2001). The putative monocarboxylate permeases of the yeast *Saccharomyces cerevisiae* do not transport monocarboxylic acids across the plasma membrane. *Yeast* **18**: 1131-43.
- McEachern, M. J. and Blackburn, E. H. (1994). A conserved sequence motif within the exceptionally diverse telomeric sequences of budding yeasts. *Proc Natl Acad Sci U S A* **91**: 3453-7.
- McEachern, M. J. and Blackburn, E. H. (1996). Cap-prevented recombination between terminal telomeric repeat arrays (telomere CPR) maintains telomeres in *Kluyveromyces lactis* lacking telomerase. *Genes Dev* **10**: 1822-34.
- McEachern, M. J. and Iyer, S. (2001). Short telomeres in yeast are highly recombinogenic. *Mol Cell* **7**: 695-704.
- McEachern, M. J., Krauskopf, A. and Blackburn, E. H. (2000). Telomeres and Their Control. *Annu Rev Genet* **34**: 331-358.
- Mefford, H. C., Linardopoulou, E., Coil, D., van den Engh, G. and Trask, B. J. (2001). Comparative sequencing of a multicopy subtelomeric region containing olfactory receptor genes reveals multiple interactions between non-homologous chromosomes. *Hum Mol Genet* **10**: 2363-72.

- Moxon, E. R., Rainey, P. B., Nowak, M. A. and Lenski, R. E. (1994). Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Biol* **4**: 24-33.
- Natarajan, S., Groff-Vindman, C. and McEachern, M. J. (2003). Factors Influencing the Recombinational Expansion and Spread of Telomeric Tandem Arrays in *Kluyveromyces lactis*. *Eukaryot Cell* **2**: 1115-27.
- Naumov, G. I., Naumova, E. S. and Korhola, M. P. (1995). Chromosomal polymorphism of *MEL* genes in some populations of *Saccharomyces cerevisiae*. *FEMS Microbiol Lett* **127**: 41-5.
- Naumov, G. I., Naumova, E. S. and Michels, C. A. (1994). Genetic variation of the repeated *MAL* loci in natural populations of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*. *Genetics* **136**: 803-12.
- Naumov, G. I., Naumova, E. S., Sancho, E. D. and Korhola, M. P. (1996). Polymeric *SUC* genes in natural populations of *Saccharomyces cerevisiae*. *FEMS Microbiol Lett* **135**: 31-5.
- Ness, F. and Aigle, M. (1995). *RTM1*: a member of a new family of telomeric repeated genes in yeast. *Genetics* **140**: 945-56.
- Neueglise, C., Feldmann, H., Bon, E., Gaillardin, C. and Casaregola, S. (2002). Genomic evolution of the long terminal repeat retrotransposons in hemiascomycetous yeasts. *Genome Res* **12**: 930-43.
- Pavlov, Y. I., Newlon, C. S. and Kunkel, T. A. (2002). Yeast origins establish a strand bias for replicational mutagenesis. *Mol Cell* **10**: 207-13.
- Pryde, F. E., Gorham, H. C. and Louis, E. J. (1997). Chromosome ends: all the same under their caps. *Curr Opin Genet Dev* **7**: 822-8.
- Pryde, F. E. and Louis, E. J. (1997). *Saccharomyces cerevisiae* telomeres. A review. *Biochemistry* **62**: 1232-41.

- Royle, N. J., Clarkson, R. E., Wong, Z. and Jeffreys, A. J. (1988). Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics* **3**: 352-60.
- Shore, D. (1998). Telomeres--unsticky ends. *Science* **281**: 1818-9.
- Teng, S. C. and Zakian, V. A. (1999). Telomere-telomere recombination is an efficient bypass pathway for telomere maintenance in *Saccharomyces cerevisiae*. *Mol Cell Biol* **19**: 8083-93.
- Teunissen, A. W. and Steensma, H. Y. (1995). Review: the dominant flocculation genes of *Saccharomyces cerevisiae* constitute a new subtelomeric gene family. *Yeast* **11**: 1001-13.
- Tracy, R. B., Baumohl, J. K. and Kowalczykowski, S. C. (1997). The preference for GT-rich DNA by the yeast Rad51 protein defines a set of universal pairing sequences. *Genes Dev* **11**: 3423-31.
- Trask, B. J., Friedman, C., Martin-Gallardo, A., Rowen, L., Akinbami, C., Blankenship, J., Collins, C., Giorgi, D., Iadonato, S., Johnson, F., Kuo, W. L., Massa, H., Morrish, T., Naylor, S., Nguyen, O. T., Rouquier, S., Smith, T., Wong, D. J., Youngblom, J. and van den Engh, G. (1998). Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum Mol Genet* **7**: 13-26.
- Tzfati, Y., Fulton, T. B., Roy, J. and Blackburn, E. H. (2000). Template boundary in a yeast telomerase specified by RNA structure. *Science* **288**: 863-7.
- van Overveld, P. G., Lemmers, R. J., Deidda, G., Sandkuijl, L., Padberg, G. W., Frants, R. R. and van der Maarel, S. M. (2000). Interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity. *Hum Mol Genet* **9**: 2879-84.

- Vergnaud, G., Mariat, D., Apiou, F., Aurias, A., Lathrop, M. and Lauthier, V. (1991). The use of synthetic tandem repeats to isolate new VNTR loci: cloning of a human hypermutable sequence. *Genomics* **11**: 135-44.
- Verstrepen, K. J., Derdelinckx, G., Verachtert, H. and Delvaux, F. R. (2003). Yeast flocculation: what brewers should know. *Appl Microbiol Biotechnol* **61**: 197-205.
- Wang, J. and Needleman, R. (1996). Removal of Mig1p binding site converts a *MAL63* constitutive mutant derived by interchromosomal gene conversion to glucose insensitivity. *Genetics* **142**: 51-63.
- Wang, R. F. and Kushner, S. R. (1991). Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*. *Gene* **100**: 195-9.
- Weilbaecher, R. G. and Lundblad, V. (1999). Assembly and regulation of telomerase. *Curr Opin Chem Biol* **3**: 573-7.
- Wellinger, R. J. and Sen, D. (1997). The DNA structures at the ends of eukaryotic chromosomes. *Eur J Cancer* **33**: 735-49.
- Wells, R. A., Green, P. and Reeders, S. T. (1989). Simultaneous genetic mapping of multiple human minisatellite sequences using DNA fingerprinting. *Genomics* **5**: 761-72.
- Wong, Z., Royle, N. J. and Jeffreys, A. J. (1990). A novel human DNA polymorphism resulting from transfer of DNA from chromosome 6 to chromosome 16. *Genomics* **7**: 222-34.
- Wray, L. V., Jr., Witte, M. M., Dickson, R. C. and Riley, M. I. (1987). Characterization of a positive regulatory gene, *LAC9*, that controls induction of the lactose-galactose regulon of *Kluyveromyces lactis*: structural and functional relationships to *GAL4* of *Saccharomyces cerevisiae*. *Mol Cell Biol* **7**: 1111-21.
- Zakian, V. A. (1995). Telomeres: beginning to understand the end. *Science* **270**: 1601-7.

Zou, S., Ke, N., Kim, J. M. and Voytas, D. F. (1996). The *Saccharomyces* retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci.

*Genes Dev* **10**: 634-45.

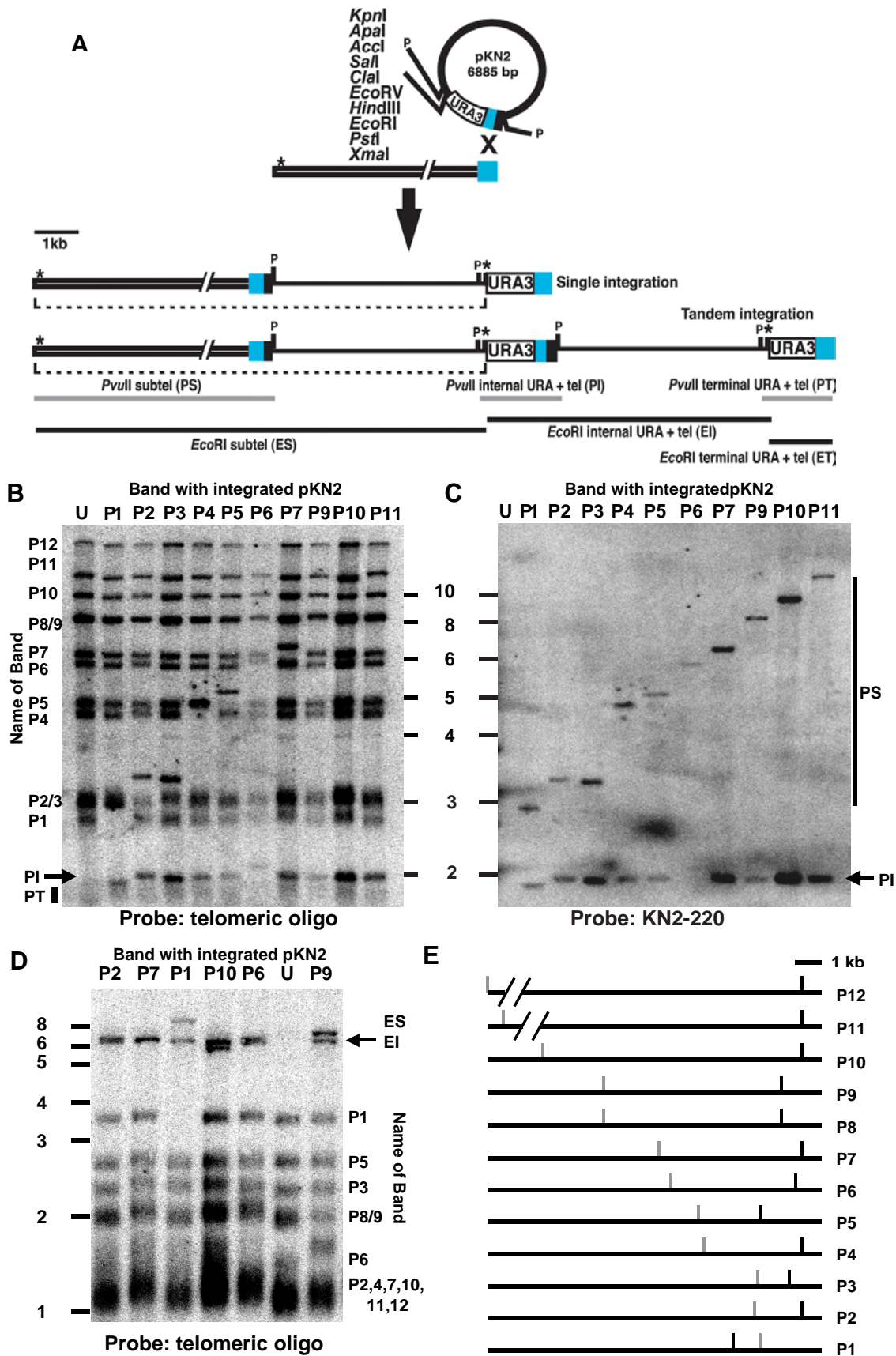
Zou, S., Wright, D. A. and Voytas, D. F. (1995). The *Saccharomyces* Ty5 retrotransposon family is associated with origins of DNA replication at the telomeres and the silent mating locus

HMR. *Proc Natl Acad Sci U S A* **92**: 920-4.

**Table 2.1: Strains used in this study.**

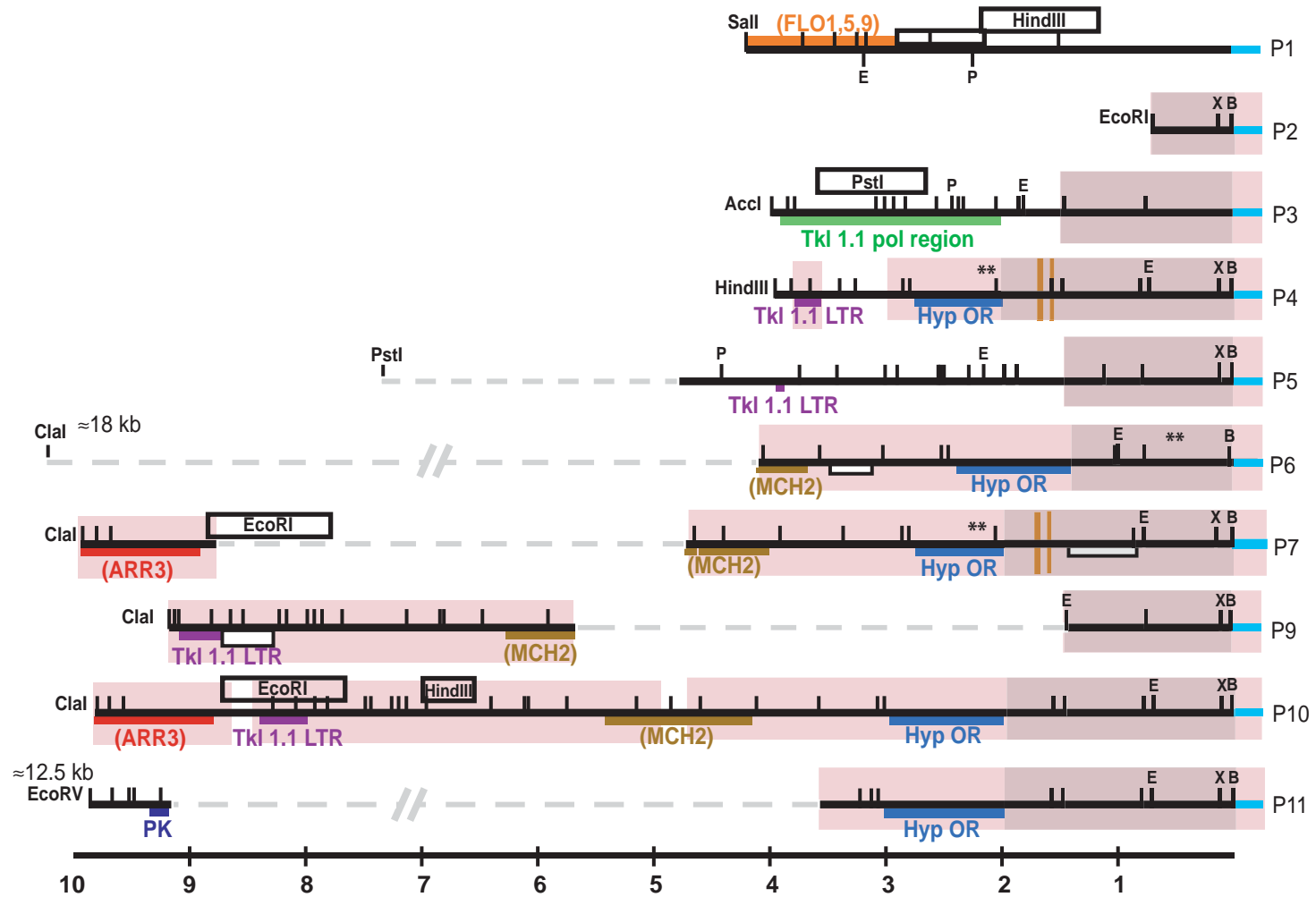
Strain	Source or reference
<i>K. lactis</i>	
7B520	(Wray et al., 1987)
GG1935	(Zonneveld and van der Zanden, 1995)
CECT 1121	From the Spanish Type Culture Collection (CECT)
CECT 10361	(Belloch et al., 2002)
<i>K. lactis var. drosophilarum</i>	
CECT 11397	
CECT 10390	
<i>K. dobzhanskii</i>	
CECT 10180	
<i>K. marxianus</i>	
CECT 10315	
CECT 1442	
<i>K. wickerhamii</i>	
CECT 1966T	

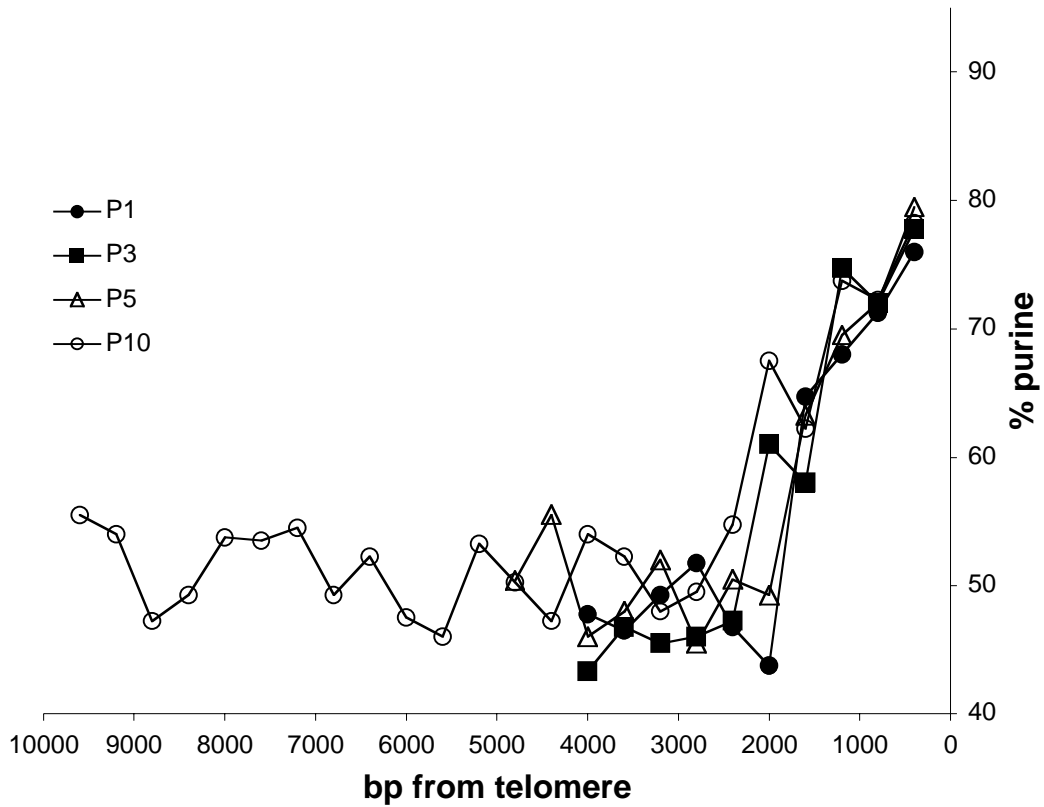
**Figure 2.1: Structure and integration of pKN2 at *K. lactis* telomeres.** A) Scheme for integration. Blue boxes represent *K. lactis* telomeric sequence, both native and integrated. The thick black bar shows the location of the vector-specific probe KN2-220 used to identify transformants. Asterisks show the locations of vector restriction sites used for cloning (including *EcoRI*), and P's represent *PvuII* sites. Integration scenarios showing single or double inserts are presented. Bracketed regions shown with dashed lines indicate fragments that would be cloned after restriction digestion and ligation. Bands from *EcoRI* and *PvuII* digests that are visible with probes in this study are indicated at the bottom of the panel and labeled in the following Southern. B) Southern blot of *PvuII*-digested genomic DNA from pKN2-transformed cells is shown hybridized to a telomeric probe. The position of the PI band is indicated by the arrow, and PT is the smear below PI. C) Filter of panel B is shown after hybridization with the KN2-220 probe (black box in panel A). D) DNA from several pKN2 transformants digested with *EcoRI* is shown hybridized with a telomeric probe. The bands visible between 6 and 8 kb are the subtelomeric bands (ES) that have been shifted by the integrated plasmid and the additional copy of the plasmid that has integrated (the lower band – EI marked by an arrow). ES and EI migrated together in lanes P2, P7, and P6. ET (~2 kb) is not labeled as it is obscured by the other bands in that size range. Lanes in all panels are labeled according to the *PvuII* telomeric fragment that contains the integrated plasmid. The lanes labeled U show DNA from an untransformed sample. Size markers are shown in kilobases. E) Shown is a diagram of the *EcoRI* (black bar) and *PvuII* (gray bar) sites closest to the telomere for each of the twelve chromosome ends of 7B520.





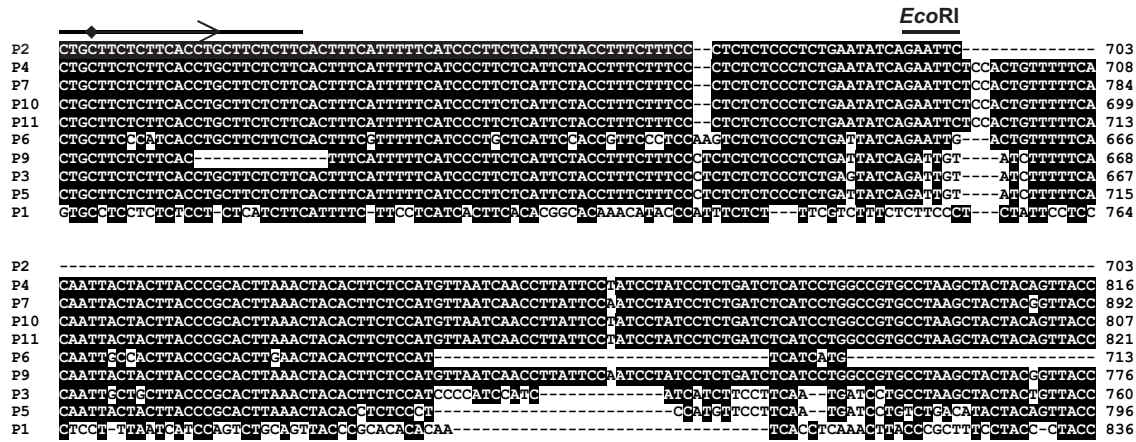
**Figure 2.2: Overall view of cloned sequence from ten *K. lactis* telomeres.** Black lines represent sequence that has been cloned and sequenced. Dashed gray lines represent sequence that has been cloned but not sequenced. Blue tips are telomeric sequences (250 bp shown). Colored features labeled with gene names in parentheses are sections homologous to the *S. cerevisiae* gene indicated. Those labeled Hyp OR are homologous to several oxidoreductases—the highest BLASTX score was from a hypothetical protein in *N. crassa*. Other ORFs that encode  $\geq 100$  aa peptides are open, unlabeled rectangles with black outline. Lightly colored rectangles in backgrounds are used to show sequences that share homology with at least one other region in this study. The darker regions of this shading next to telomeres represent locations of the R element described in text. Double asterisks on P4, P7, and P6 show where gaps exist in sequence relative to other subtelomeric regions (250 bp in P4 and P7 and 480 bp in P6), and vertical orange bars on P4 and P7 show regions of poor alignment to other subtelomeres. Black vertical bars represent restriction sites (unmarked – *Sau3AI*, B= *BsrBI*, X= *XbaI*, E= *EcoRI*, P= *PvuII*). The left-most restriction enzyme on each end is the enzyme used to clone that particular end. On P1, P3, P7, and P10, probes used in hybridization studies are shown with open rectangles labeled with restriction enzyme used in subcloning (see Table 2.2).



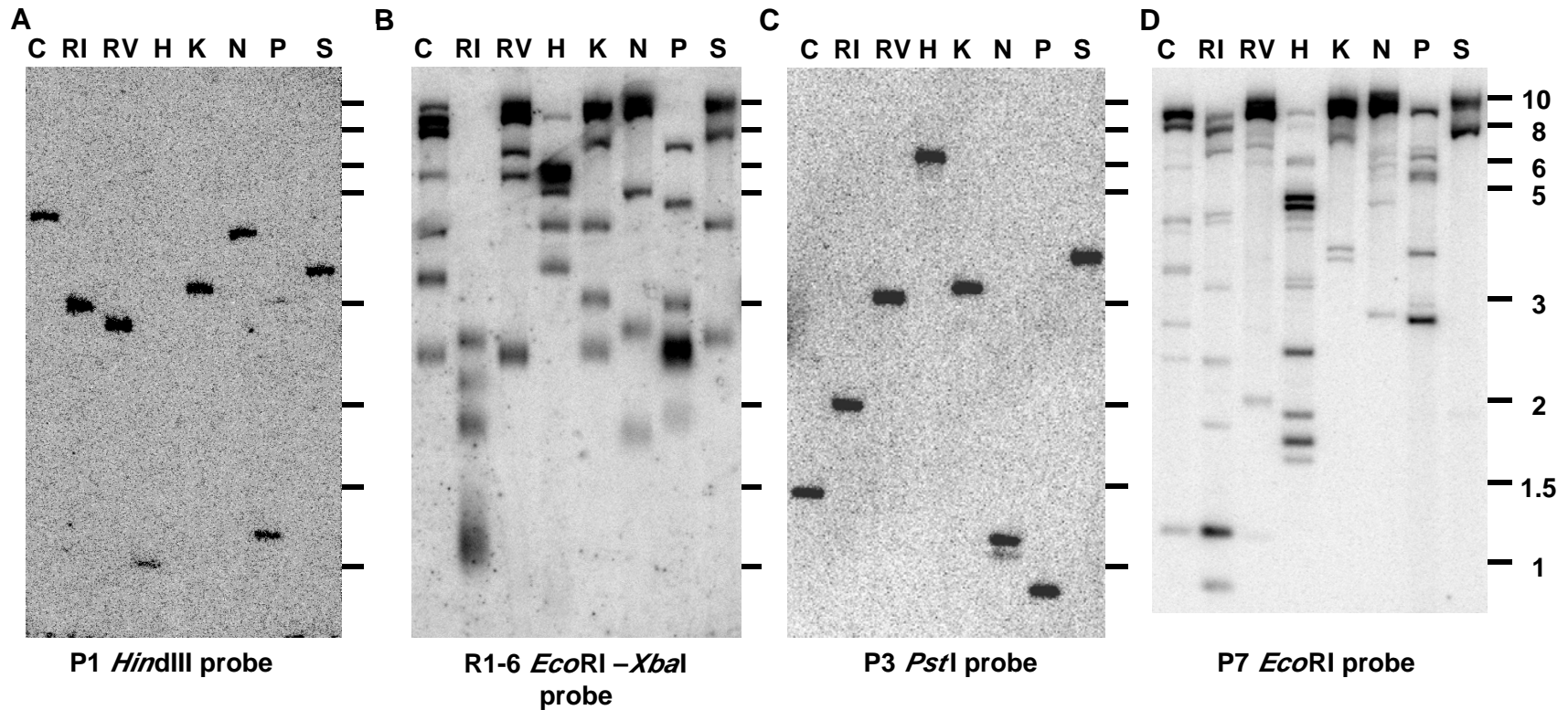


**Figure 2.3: Purine strand bias in the immediate subtelomeric regions of *K. lactis*.** Each point shows (G + A)% for the preceding 400-bp interval. The zero position indicates the junction of telomeric and subtelomeric sequences. Subtelomeric regions from four telomeres that contain some potentially unique DNA are shown.

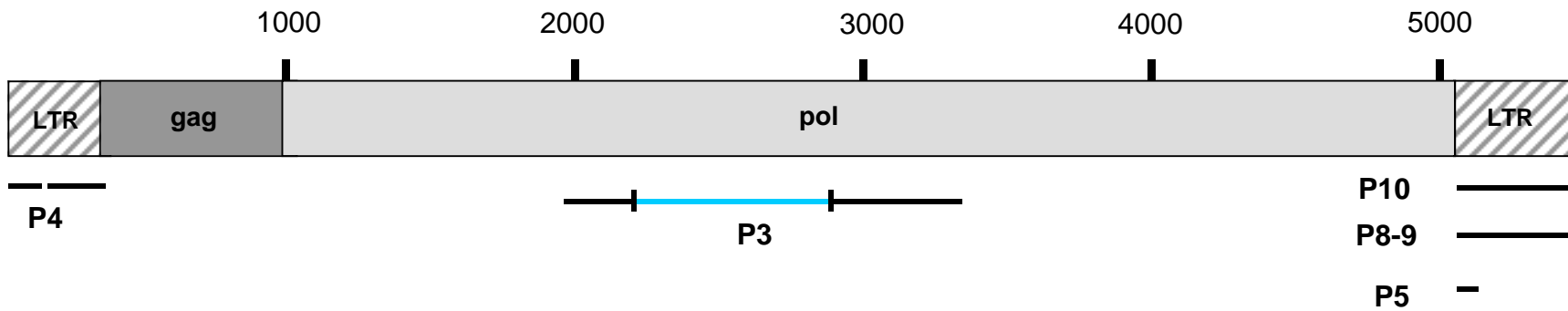




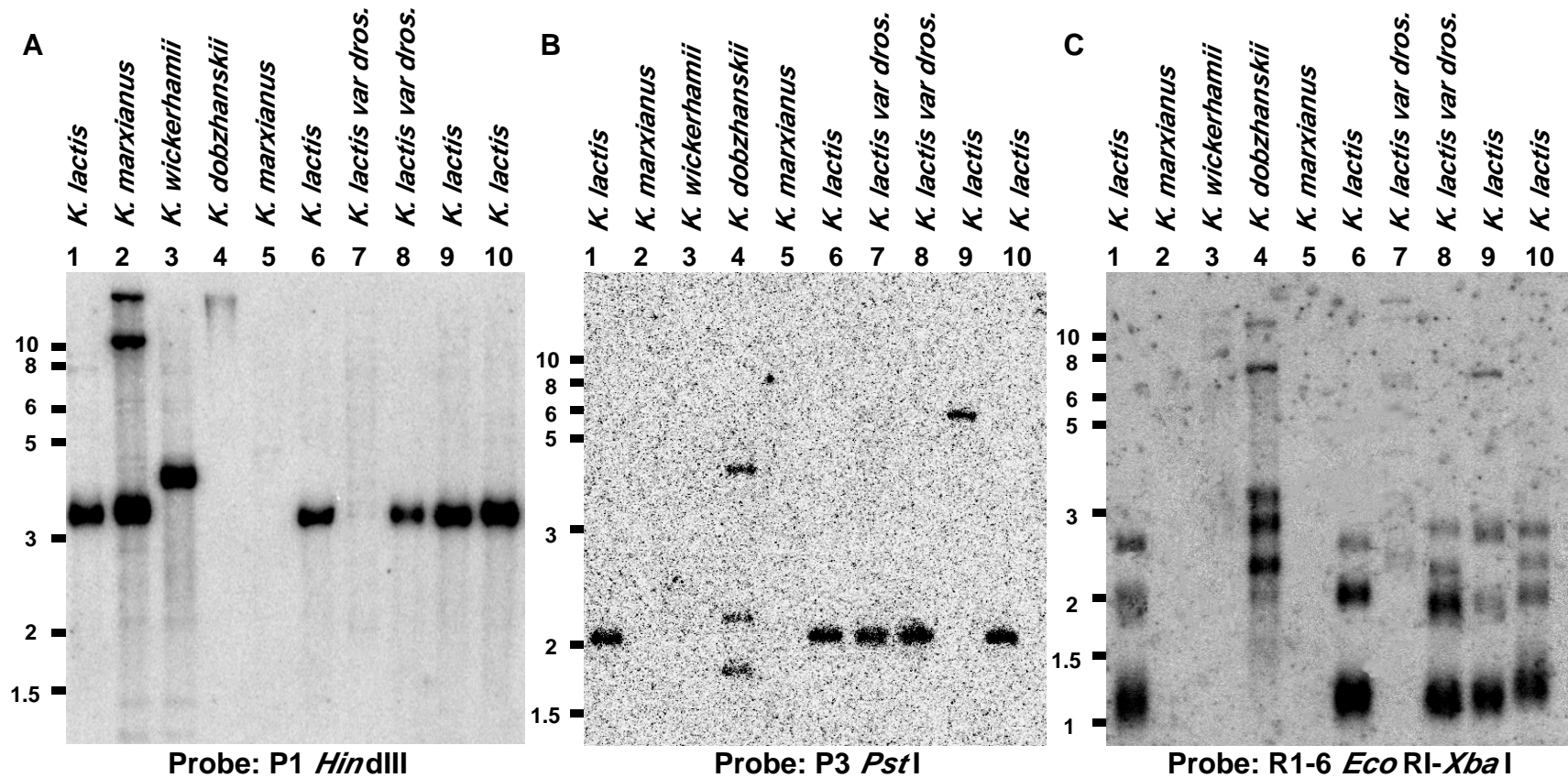
**Figure 2.4: Alignment of immediately subtelomeric sequences from ten *K. lactis* chromosome ends.** This alignment includes 50 bp of telomeric sequence (upper left) and about 800 bp just internal to the telomere. Conserved bases are highlighted black. Features are labeled above. Sequences are in order according to *EcoRI* restriction pattern (see Figure 2.1). The available P2 sequence ends at the *EcoRI* site, and P6 has a large deletion beginning at position 713 that extends beyond the end point shown.



**Figure 2.5: Hybridization of subtelomeric probes to 7B520 genomic DNA.** Shown are Southern blots of 7B520 DNA digested with eight different restriction enzymes and hybridized to four different subtelomeric probes as indicated. Probes are mapped in Figure 2.2. C, *ClaI*; RI, *EcoRI*; RV, *EcoRV*; H, *HindIII*; K, *KpnI*; N, *NsiI*; P, *PstI*; S, *SalI*.



**Figure 2.6: Retrotransposon Tk1.1 homology in 7B520 subtelomeric regions.** Tk1.1 is labeled according to GenBank annotations (Accession number AJ439548). Regions of Tk1.1 that are present in different subtelomeric regions are indicated as labeled black bars. Marked region in the P3 fragment shows the location of the *Pst*I probe used in this study (Figure 2.2). See text for details of the alignments.



**Figure 2.7: Hybridization of subtelomeric probes to other *Kluyveromyces* yeasts.** Southern blots show genomic DNA from several *Kluyveromyces* yeasts digested with *Eco*RI hybridized to the subtelomeric probes discussed in this study. Three of these hybridizations are shown above. Strains in each lane are as follows: 1, CECT 1121; 2, CECT 1442; 3, CECT 1966T; 4, CECT 10180; 5, CECT 10315; 6, CECT 10361; 7, CECT 10390; 8, CECT 11397; 9, GG1935; 10, 7B520



## CHAPTER 3

### CONCLUSIONS AND FUTURE DIRECTIONS

Subtelomeric regions of several organisms are subjects of investigation because of the generally dynamic structure of these regions, the adaptive importance genes sometimes found there, and their various effects on chromosomes, particularly on telomeres. Knowledge from a number of species may facilitate the proposal of models of how subtelomeric ends adapted to become so potentially versatile yet specific for each species.

There are a number of questions that can be addressed and projects that can be pursued using the *K. lactis* sequences reported here. One step would be to complete the sequencing and characterization of the available cloned subtelomeric sequences for 7B520. This would include cloning the other two telomere regions and pulsed-field gel analysis to identify the chromosome arm on which each subtelomere is located.

The genome of a different strain of *K. lactis* (CBS 2359) is currently being sequenced and annotated (Ozier-Kalogeropoulos et al., 1998; Feldmann, 2000). The information from our work could be of use to this genome project as telomeric fragments of a library might be difficult to assign to a particular chromosome end due to the extensive subtelomeric homology. With the degree of similarity that appears to exist between strains, the sequence from these marked telomere transformants would be useful in organizing the end sequences. If the variation is such that sequence comparison does not resolve all ends, this sequenced strain's ends may have to be cloned as was done in our study. The sequences from both strains can then be used to get a first look at the extent of variation among *K. lactis* subtelomeres.

Also, subtelomeric sequences from 7B520 and CBS 2359 along with subtelomeric sequences that are available from other species, including *S. cerevisiae* and other *Kluyveromyces* yeasts, can be used in comparing many “related” subtelomeric regions in variation and copy number of subtelomeric genes and other elements. Having such closely related sequences to examine is already being used to study genomic evolution. A study of the plastic subtelomeric regions of these organisms is clearly an interesting one to pursue.

Another potential line of investigation using subtelomeric sequences from the 7B520 strain of *K. lactis* is examination of the subtelomeric recombination that occurs in various telomerase-template mutants in this strain. Previously, highly elevated recombination rates have been measured in short-telomere mutants using loss of a *URA3* gene inserted about 120 bp from a telomere as the measurement and observed by loss of *EcoRI* bands in the restriction pattern (McEachern and Iyer, 2001). These experiments give only information regarding genetic exchange in the regions largely less than 1 kb from the telomere. The subtelomeric sequences of 7B520 can be used in studies to ascertain the extent of this exchange and provide clues about mechanisms by which it occurs..

Two models of such exchange are localized non-reciprocal exchange and break-induced replication (Figure 3-1). Localized non-reciprocal exchange refers to gene conversion for which the end of the exchange occurs before the end of the chromosome is reached. The replaced sequence therefore constitutes a relatively small region. Break-induced replication is when a broken end invades another chromosome to use as a template, and the template strand is copied until its end potentially resulting in the copy of a very long sequence (Malkova et al., 1996). Evidence is accumulating that these two categories of events are different in how they arise.

To resolve these two models experimentally in *K. lactis*, a telomeric tract composed of only mutant repeats can aid as a marker in discerning between these two models (Underwood and McEachern, 2001). If internal sequence, or an internal marker gene such as *URA3*, connected to a marked mutant telomere has been lost or duplicated in conjunction with loss or duplication of the marked telomere, the mechanism of this exchange is break-induced replication as it extends from the internal point to the end of the chromosome (Figure 3-1). In *S. cerevisiae*, recombinational repair of chromosomal breaks in subtelomeric regions was break-induced replication in 75% of cases observed (Ricchetti et al., 2003). In *K. lactis*, initial studies have been done using an assay placing the *S. cerevisiae URA3* cassette 120 bp upstream of the mutant telomere. When loss of *URA3* was selected, the mutant telomeric repeats were lost as well indicating that loss of *URA3* resulted from a mechanism like BIR (Natarajan, 2002). However, this only gives information regarding sequence less than 1.2 kb from the telomere. The available sequence would enable a more exhaustive study of this issue by studying positions further from the telomeres. This would contribute to the study of recombinational telomere maintenance in the absence of telomerase in *K. lactis*.

*K. lactis* subtelomeric sequence, especially potential coding regions, may be useful in exploring adaptation and change in these regions. Determining functions of putative genes encoded there would facilitate using them to examine adaptability in these regions. Two putative genes in these regions, *FLO1* and *ARR3*, have assigned functions in *S. cerevisiae* (flocculation and arsenite resistance, respectively); this gives a starting point in choosing environments that would favor these genes in order to monitor subtelomeric adaptations. Flocculation is a desirable feature in brewing yeasts because the cell adhesion results in clear separation of yeast from the media, and industrial strains are known to have different flocculation profiles due to genetic

variability (Verstrepen et al., 2003). Putting 7B520 under selection for these genes could be useful in providing examples in the formation and alteration of subtelomeric gene families.

Along the same lines, experiments can be designed that would reveal any bias in subtelomeric gene conversion. This may involve investigation using the compositional bias found in these regions. One possibility for doing this would be to collaborate with a lab studying recombination *in vitro* and examine whether *K. lactis* R element sequences are better at strand invasion, using the purine-rich strand, than other sequences including the R element's pyrimidine-rich strand. Recombination studies involving bias have shown that gene conversion could favor GC base pairs over AT (Marais, 2003) and that Rad51p, the yeast homolog of the RecA protein, has a preference for GT-rich DNA (Tracy et al., 1997). These considerations cannot be applied to our sequence at this time, but they may be used in designing experiments or models in studying subtelomeric gene conversion on this level.

It has been suggested, using sequence data, that in *S. cerevisiae* replication-associated mutational pressure is more easily observed in subtelomeric regions due to the fixed roles of leading and lagging strands (Gierlik et al., 2000). More recently, it has been established that strand bias in replicational mutagenesis associated with active replication origins occurs in *S. cerevisiae* (Pavlov et al., 2002). It would be useful to test whether the *K. lactis* R element is subject to such biased mutation and if the border of the R element occurs at an ARS element. Efforts are underway to identify subtelomeric ARS elements in this *K. lactis* sequence. If present, a similar approach to that of *S. cerevisiae* may be used to study such strand bias with *K. lactis* ARSes.

Transcriptional silencing at native telomeres has been demonstrated to occur in *K. lactis* (Gurevich et al., 2003). Since there are ORFs located close enough to telomeres to be candidates

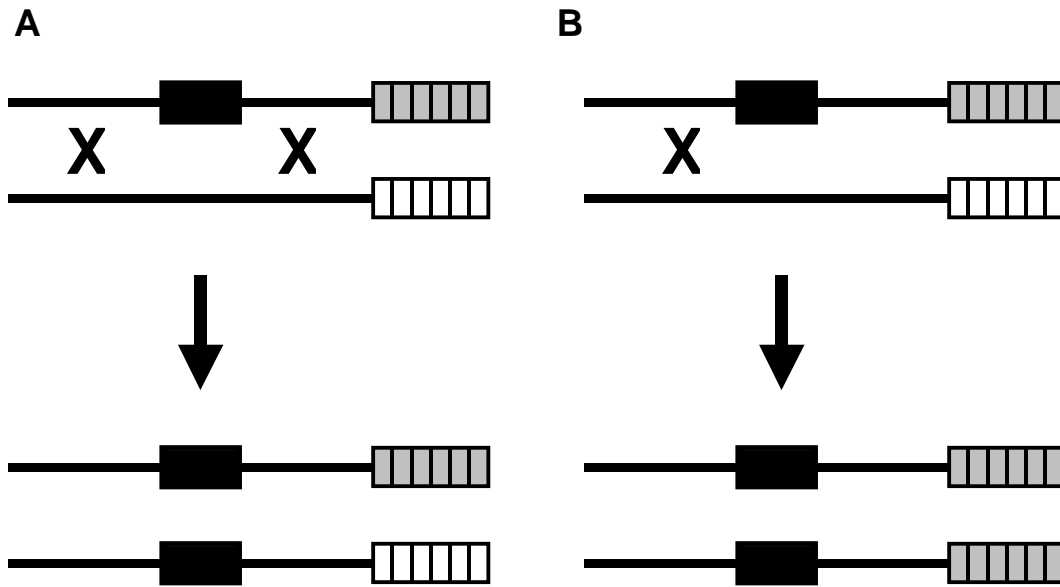
for silenced genes, our subtelomeric sequence can aid in this line of investigation as well. If *K. lactis* subtelomeric ORFs prove to encode functional proteins, they would be very interesting to examine whether natural silencing occurs at these genes.

The subtelomere cloning project described in this work will serve mainly as a tool in recombinational studies. Yet, the information it will provide to the field of subtelomeric study cannot be foreseen fully at this time. Once this sequence undergoes more intense bioinformatic analysis and comparison with appropriate sequences, there will likely be much more information gained from the results of this labor than is reported here.

## References

- Feldmann, H. (2000). Genolevures--a novel approach to 'evolutionary genomics'. *FEBS Lett* **487**: 1-2.
- Gierlik, A., Kowalczyk, M., Mackiewicz, P., Dudek, M. R. and Cebrat, S. (2000). Is there replication-associated mutational pressure in the *Saccharomyces cerevisiae* genome? *J Theor Biol* **202**: 305-14.
- Gurevich, R., Smolikov, S., Maddar, H. and Krauskopf, A. (2003). Mutant telomeres inhibit transcriptional silencing at native telomeres of the yeast *Kluyveromyces lactis*. *Mol Genet Genomics* **268**: 729-38.
- Malkova, A., Ivanov, E. L. and Haber, J. E. (1996). Double-strand break repair in the absence of *RAD51* in yeast: a possible role for break-induced DNA replication. *Proc Natl Acad Sci U S A* **93**: 7131-6.
- Marais, G. (2003). Biased gene conversion: implications for genome and sex evolution. *Trends Genet* **19**: 330-8.

- McEachern, M. J. and Iyer, S. (2001). Short telomeres in yeast are highly recombinogenic. *Mol Cell* **7**: 695-704.
- Natarajan, S. (2002). Recombinational Telomere Elongation in *Kluyveromyces lactis*, *Genetics*, University of Georgia, pp. 177.
- Ozier-Kalogeropoulos, O., Malpertuy, A., Boyer, J., Tekaiia, F. and Dujon, B. (1998). Random exploration of the *Kluyveromyces lactis* genome and comparison with that of *Saccharomyces cerevisiae*. *Nucleic Acids Res* **26**: 5511-24.
- Pavlov, Y. I., Newlon, C. S. and Kunkel, T. A. (2002). Yeast origins establish a strand bias for replicational mutagenesis. *Mol Cell* **10**: 207-13.
- Ricchetti, M., Dujon, B. and Fairhead, C. (2003). Distance from the chromosome end determines the efficiency of double strand break repair in subtelomeres of haploid yeast. *J Mol Biol* **328**: 847-62.
- Tracy, R. B., Baumohl, J. K. and Kowalczykowski, S. C. (1997). The preference for GT-rich DNA by the yeast Rad51 protein defines a set of universal pairing sequences. *Genes Dev* **11**: 3423-31.
- Underwood, D. H. and McEachern, M. J. (2001). Totally mutant telomeres: single-step mutagenesis of tandem repeat DNA sequences. *Biotechniques* **30**: 934-5, 938.
- Verstrepen, K. J., Derdelinckx, G., Verachtert, H. and Delvaux, F. R. (2003). Yeast flocculation: what brewers should know. *Appl Microbiol Biotechnol* **61**: 197-205.



**Figure 3-1: Possible Models of Subtelomeric Gene Conversion.** The two models discussed and how they would be discerned from one another are shown above. Segmented blocks on the ends represent telomeres (white = wild-type and gray = mutant repeats). Solid black lines represent subtelomeric sequence with the black box representing a subtelomeric marker gene. (A) Localized subtelomeric gene conversion: spread of a subtelomeric marker not linked to spread of the marked telomere. (B) Subtelomeric gene conversion via break-induced replication: spread of the subtelomeric marker and marked telomeres are linked.

## APPENDIX A

### pKN2 TRANSFORMANTS AND DERIVED PLASMIDS

Transformants in this study were assigned long term names based on the order in which they were picked from transformation plates. For example, the seventeenth transformant picked has been named MT17 (marked telomere strain 17). A plasmid derived from a transformant is named using three identifiers: the transformant from which it is derived, its number in the screen of plasmids, and the restriction enzyme used to rescue the plasmid. For example, the plasmid 17-7 *SalI* was derived from MT17, the seventh plasmid prep of that particular plasmid screen, and rescued using *SalI*. An exception is that any plasmids rescued by *EcoRI* do not have the enzyme name incorporated into its name in any of the records.

Some plasmids were used as sources of fragments for use as subtelomeric probes in this study. To facilitate the long-term use of these probes, the fragments used as probes were subcloned into the high copy plasmid pBLUESCRIPT SK<sup>®</sup>. The name of such a plasmid was derived from the transformant used and the restriction enzyme(s) used in subcloning the fragment. The plasmid p17H, for example, contains a *HindIII* fragment from the clone 17-7 *SalI*.

Table A.1 lists all plasmids used in this study. The *PvuII* band number was how we named the chromosome ends; this was described in more detail in the Results. The sizes given for the pKN2-based plasmids refer to the sizes of the subtelomeric inserts (total *K. lactis* insert size minus telomeric DNA). Exact numbers are based on sequence recovered (Figure 2.2), and the approximated sizes are based on Southern data (not shown).



Two of the plasmids are not pure in that additional restriction fragments were ligated into the resulting clone. These are indicated by the asterisks in Table A.1. 17-13 has one additional *EcoRI* fragment, and 43-1 *PstI* has two additional *PstI* fragments. Sequence from 17-7 *SalI* was used to confirm that the extra fragment in 17-13 was not contiguous to the terminal *EcoRI* fragment (i.e. not the result of a partial *EcoRI* digestion). No such verification can be made for the 43-1 *PstI* plasmid at this time, but only the sequence distal to the terminal *PstI* restriction site was used in this study and in estimating size of insert in Table A.1.

**Table A.1: pKN2 transformants and plasmids used in this study.**

<i>PvuII</i> band	Transformant number	<i>EcoRI</i> recovered plasmid + size	Large recovered plasmid and enzyme used + size	Plasmids with fragments for specific probes and size of probe
1	17	17-13* 2598 bp	17-7 <i>SalI</i> 4160 bp	p17H – <i>HindIII</i> fragment 1012 bp
2	15	15-11 703 bp	N/A	
3	10	10-5 1790 bp	10-28 <i>AccI</i> 3959 bp	p10Pst – <i>PstI</i> fragment 932 bp
4	6	6-5 694 bp	6-60 <i>HindIII</i> 3930 bp	
5	43	43-31 2151 bp	43-1 <i>PstI</i> * ~ 7000 bp	
6	19	19-36 952 bp	19-32 <i>ClaI</i> > 12 kb	
7	16	16-4 770 bp	16-84 <i>ClaI</i> ~ 10 kb	p16RI – <i>EcoRI</i> fragment 1057 bp
8/9	61	61-13 1423 bp	61-54 <i>ClaI</i> ~ 9 kb	
10	18	18-16 685 bp	18-6 <i>ClaI</i> 9815 bp	p18RI – <i>EcoRI</i> fragment 1057 bp p18H – <i>HindIII</i> fragment 440 bp
11	14	14-10 699 bp	14-20 <i>EcoRV</i> ~ 12 kb	