THE ANALYSIS OF VARIATION AT THE HUMAN Xp:Yp TELOMERE

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Duncan Martin Baird
Department of Genetics
University of Leicester

January 1996

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346


LEICESTER UNIVERSITY
LIBRARY

In memory of Richard Evans 'Rich'

CONTENTS

THE ANALYSIS OF VARIATION AT THE HUMAN Xp:Yp TELOMERE

By Duncan Martin Baird

ABSTRACT

The DNA immediately flanking the human Xp:Yp telomere exhibits a high level of sequence polymorphism and strong linkage disequilibrium, resulting in a limited number of highly diverged haplotypes. The sequence divergence suggests that these haplotypes are ancient. Orthologous sequences in chimpanzees and gorillas are more diverged than between synonymous sequences at other loci. Balancing selection may have contributed to the maintenance of the highly diverged haplotypes flanking the human Xp:Yp telomere and may explain the high sequence divergence in closely related species. A system has been developed to assay the distribution of telomere and variant repeats within the proximal 120 repeats of Xp:Yp telomeres to create a telomere map. The distributions of these repeats is highly polymorphic with estimated heterozygosities in excess of 99%. The mutation rate underlying this variation was measured directly as 6.25×10^{-3} per gamete. Alleles were grouped by similarities in their telomere maps, these groups of alleles also share the same haplotype in the telomere flanking DNA. This suggests that these alleles have evolved along haploid lineages and that the predominant mutational mechanisms influencing the evolution of sequences at the proximal end of this telomere must be of an intra-allelic nature. Comparison of alleles suggests that most of the differences could be accounted for by small localised replication-slippage like events. Analysis of the chimpanzee and gorilla sequences orthologous to the human Xp:Yp telomere revealed that a telomere was not present at the same location in these species. Instead, two small interstitial blocks of telomere-like repeats were present in gorilla and an array of chimpanzee and gorilla specific subterminal satellite was present in chimpanzees. Therefore, the location of the Xp:Yp telomere is unique to the human lineage and may therefore be relatively new in evolutionary terms.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my Biology teacher Mr. Roger Poland for his extraordinary enthusiasm for all things Biological and to Dr. Graham Taylor for introducing me to Human Genetics in an essentially lower eukaryotic/prokaryotic Genetics degree. Thanks go to my parents and family for supporting me through out my education and for not questioning too much my choice of career.

Thanks go to all members of G19 past and present for making my time so enjoyable and for setting such high standards, these include Annette, Darren, Keiji, John Armour, Jane, Yuri, Chong-Lek, Hong Lang, Andy, Rita, Jen, Tara, Esther, Max, John Murray, Jo, Celia, Shaojie, Jackie, Moira, David, Philippe and Neale. Special thanks go to Ila for the smooth running of the Laboratory, to John Armour for advice on every thing, Yuri for help with statistics and to David, Philippe, Andy, Mark and Darren for discussions on all things repetitive. Credit also goes to those who make this department run so smoothly, these include Terry, Margaret, Sandy, Jenny Whiteman, Jenny Foxon and Sarah. Thanks also go the Katherine Lilley, Elli, Beth and Stewart for their help with the ABI machine, also to John Keyte for the emergency oligo service.

For out of lab activities I thank; David and Helen for pinball and repetitive beats, Robin for steaks and beer, R.J. and Bean for outdoor activities, all the members of Annettes lab for parties, and to Paul, Karl, Jen, Dean, Julie, Richard, Neale and Philippe for assorted other leisure activities. Of course very special thanks go to Katherine for making my time in Leicester so good and for proof reading the unreadable.

Finally, I thank Alec for allowing me to work in this excellent laboratory and of course thanks go to Nicola for her guidance and support during this work. I am sure this will be the first of many theses to come from her expanding group.

This work was supported by a grant from the MRC.

PUBLICATIONS

Some of this work has been published previously:

Baird, D. M., Jeffreys, A. J. and Royle, N. J. (1994). Extreme variability in the Xp:Yp pseudoautosomal telomere revealed by telomere variant repeat mapping. *Am. J. Hum. Genet.* 35 **[Suppl]**: A123.

Baird, D. M., Jeffreys, A. J. and Royle, N. J. (1995). Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human Xp/Yp telomere. *EMBO J.* 14: 5433-5443.

Some of this work is the subject of patent applications.

ABBREVIATIONS

A	adenine
ARMS	amplification refractory mutation system
ATP	adenosine 5′-triphosphate
bp, kb, Mb	base pair; kilo-, mega-
BSA	bovine serum albumin
C	cytosine
cDNA	complementary deoxyribonucleic acid
CEPH	Centre d'Etude du Polymorphism Humain
dATP	2′-deoxyadenosine 5′-triphosphate
dCTP	2′-deoxycytosine 5′-triphosphate
dGTP	2′-deoxyguanosine 5′-triphosphate
DNA	deoxyribonucleic acid
dTTP	2′-deoxythymidine 5′-triphosphate
EDTA	ethylenediaminetetra-acetic acid
FISH	fluorescence *in situ* hybridisation
G	guanine
g, mg, µg, ng	grams; milli-, micro-, nano
HCl	hydrochloric acid
HLA	human leukocyte antigens
HNPCC	hereditary non-polyposis colon cancer
HVR	hypervariable region
ins/del	insertion/deletion
l, ml, µl	litre; milli-, micro-
LINE	long interspersed nuclear element
µ	mutation rate
M, mM, µM, nM	molar; milli-, micro-, nano-
MAT	mating type locus
MHC	major histocompatibility complex
min	minute
MS	minisatellite
mtDNA	mitochondrial DNA
MVR-PCR	minisatellite variant repeat mapping by PCR
Myr	million years
NIH	national institutes for health (USA)
nt	nucleotide
PCR	polymerase chain eeaction
PEG	polyethylene glycol

pers. comm.	personal communication
PFGE	pulsed field gel electrophoresis
RER	replication error repair
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
r p m	revolutions per minute
SDS	sodium dodecyl sulphate
sec	second
SINE	short interspersed nuclear element
SSC	saline sodium citrate
SSCP	single stranded conformational polymorphism
STR	simple tandem repeat
T	thymidine
TBE	tris-borate EDTA
THE	transposon-like human element
Tris	Tris-(hydroxymethyl)-methylamine[2-amino-(2-hydroxymethyl)-propan-1,3-diol]
TVR-PCR	telomere variant repeat mapping by PCR
USCE	unequal sister chromatid exchange
UV	ultra violet
VNTR	variable number tandem repeats
YAC	yeast artificial chromosome

Chapter 1

INTRODUCTION

Telomeres

The term telomere was coined by Hermann Muller to describe a specialised hypothetical structure required for the maintenance of chromosome stability (Muller, 1938, 1940). Barbara McClintock observed in maize embryos that the sister chromatids from a replicated broken chromosome fused, producing a dicentric chromosome. At mitosis the dicentric chromosome formed an anaphase bridge and broke (McClintock, 1938). This process occurred in subsequent cell divisions and has become known as the breakage-fusion-bridge cycle. The cycle stopped upon fertilisation of the embryo, as the broken chromosome ends lost their reactivity and the chromosomes were said to be healed (McClintock, 1939).

The structure of telomeres

Telomeres have been characterised at the molecular level in a number of eukaryotic organisms. The extreme terminus of many organisms consists of tandemly repeated simple sequence with one G-rich strand specifically oriented 5′ to 3′ towards the end of the chromosome. This sequence is conserved in many species but the length of the repeat array varies dramatically (Table 1.1.). The length of telomere repeat arrays can be tightly controlled such as those in *Euplotes* which consist of 42 TTTTGGGG repeats (Klobutcher *et al.*, 1981), or more loosely controlled such as the telomeres of humans which consist of the repeat sequence TTAGGG in heterogeneous arrays of between 5 to 20kb (Moyzis *et al.*, 1988; de Lange *et al.*, 1990; Hastie *et al.*, 1990). The extreme ends of the telomere repeat arrays of the ciliates *Oxytricha*, *Stylonychia*, *Euplotes*, *Tetrahymena*, the slime mould *Didymium* and the yeast *Saccharomyces cerevisiae* are not blunt ended, but have a 3′ overhang of 12 or more nucleotides (Klobutcher *et al.*, 1981; Henderson and Blackburn, 1989; Wellinger *et al.*, 1993; Vermeesch and Price, 1994). The single stranded overhang may be a general feature of all telomeres that consist of a tandemly repeated simple sequence.

Not all the species studied have short tandemly repeated sequences at the extreme end of the chromosomes. Attempts have been made to detect G-rich telomere repeats in *Drosophila* and other Diptera and some Coleoptera, however, no such repeats have been identified (Richards and Ausubel, 1988; Meyne *et al.*, 1989;

Table 1.1. Comparison of telomere repeat unit and array sizes in eukaryotic organisms.

Species	Telomere repeat unit sequence	Telomere repeat array size	References
Lower Eukaryotes			
Euplotes	TTTTGGGG	42 repeats	(Klobutcher *et al.*, 1981)
Tetrahymena	TTGGGG	50 repeats	(Blackburn and Gall, 1978; Larson *et al.*, 1987)
Plasmodium berghei	TT(T/C)AGGG		(Ponzi *et al.*, 1985)
Saccharomyces cerevisiae	$TG_{2-3}(TG)_{1-6}$	300 bp	(Shampay *et al.*, 1984)
Saccharomyces castellii	TCTGGGTG interspersed with $TCTGGG(TG)_{2-4}$		(Cohn and Blackburn, 1995)
Kluyvermyces lactis	ACACCACATACCT AATCAAATCCGA	250-500bp	(McEachern and Blackburn, 1994)
Candida albicans	ACGGATGTCTAA CTTCTTGGTGT	15-30 repeats	(McEachern and Hicks, 1993)
Higher Eukaryotes			
Arabidopsis thaliana	TTTAG(G/A)G	2-5 kb	(Richards and Ausubel, 1988)
Lycopersicon esculentum	TTTAGGG	20-50 kb	(Ganal *et al.*, 1991)
Nicotiana tabacum	TTTAGGG	60-160 kb	(Fajkus *et al.*, 1995)
Myrmecia formicidae	TTAGG		(Meyne *et al.*, 1995)
Bombyx mori	TTAGG		(Okazaki *et al.*, 1993)
Homo sapiens	TTAGGG	5-20 kb	(Moyzis *et al.*, 1988; de Lange *et al.*, 1990; Hastie *et al.*, 1990)
Mus carali	TTAGGG	20-30 kb	(Meyne *et al.*, 1989; Kipling and Cooke, 1990; Starling *et al.*, 1990)
Mus spretus	TTAGGG	5-20 kb	
Mus musculus strains			
C57BL/6J	TTAGGG	upto 150 kb	
DBA/2	TTAGGG	upto 150 kb	

Okazaki *et al.*, 1993). Also, viable terminal deletions of *Drosophila* chromosomes have been detected which progressively loose terminal sequences and do not terminate in G-rich telomeric sequences (Biessmann and Mason, 1988; Levis, 1989). This is in contrast to that observed in other eukaryotic chromosomes with simple repeated sequences at the telomeres, where the broken chromosomes are healed by the addition of telomeric repeats (Wilkie *et al.*, 1990). This suggested that the organisation of the telomeres of *Drosophila* and some other insects may be very different. Subsequently it was shown that the ends of *Drosophila* chromosomes carry one or more LINE-like retrotransposable elements (Levis *et al.*, 1993). Two families of these *Drosophila* telomeric elements have been identified, telomere associated retrotransposon (TART) and the HeT-A elements both of which are preferentially associated with the ends of chromosomes (Levis *et al.*, 1993).

The structure of subterminal regions

Repeated simple sequences have been identified at the extreme ends of many eukaryotic chromosomes. In addition to these sequences, other species specific elements have been identified adjacent to telomere repeat arrays. Many different terms have been used to identify these elements (de Lange *et al.*, 1990; Brown *et al.*, 1990; Cross *et al.*, 1990; Royle *et al.*, 1992), in this thesis they will be referred to as subterminal sequences. Subterminal sequences are of a repetitive nature, often several different classes are observed within one species. One of the main reasons for the isolation of subterminal sequences in many organisms is to generate terminal markers which can be physically mapped in respect to the terminus, so providing anchor points which will lead to the closure of genetic and physical maps.

Yeast. The best characterised subterminal repeats are the Y′ elements of *Saccharomyces cerevisiae*. These elements fall into two major size classes of 5.2 and 6.7kb and are found at half to two thirds of chromosome ends in highly diverged strains of yeast (Jager and Philippsen, 1989; Louis and Haber, 1990). Upto four Y′ elements have been detected, with a short stretch of yeast telomere repeats between adjacent elements (Chan and Tye, 1983; Walmsley *et al.*, 1984). The Y′ elements contain open reading frames encoding proteins with no strong homology to any known protein (Louis and Haber, 1992). Sequence variation between Y′ elements exists, resulting in Y′ elements within strains upto 1.13% diverged and upto 1.75% diverged between strains, the majority of this variation disrupts the open reading frames (Louis and Haber, 1992). The sequence of these elements indicates that they were derived from mobile elements, however, it appears that now transposition between chromosome ends is by mitotic and

meiotic exchanges (Horowitz *et al.*, 1984; Louis and Haber, 1990, 1992). Other subterminal elements have been identified in yeast, these are always located proximal to the Y′ elements and include the X family of sequences (Chan and Tye, 1983). Chromosomes that lack Y′ elements have been identified (Zakian and Pluta, 1989) and a fully functional yeast chromosome III has been created that lacks Y′ elements (Murray and Szostak, 1986). This suggests that the subterminal repeats identified in yeast, are not required in yeast cells for normal telomere maintenance. However, the presence of these elements may have a secondary function in providing a buffer between the transcribed regions and the telomere. This could be useful for two reasons; firstly, in the event of telomere repeat array shortening such a buffer would the delay the deletion of the transcribed sequences and secondly, the buffering effect could avoid the transcriptional silencing effect of yeast telomeres (Gottschling *et al.*, 1990).

The proterminal regions of human chromosomes. The distal regions of human chromosomes have been termed the proterminal regions. These regions have been shown to be the site of the initiation of homologous chromosome pairing (Wallace and Hulten, 1985). Chromosome specific portions of the proterminal regions may therefore be involved in homology searches between chromosomes. Homology searches could involve the synapsis of chromosomes following checks for sequence homology by single stranded nick or D-loop invasion of one chromosome by another (Carpenter, 1979). If the homology search is successful, the outcome of these heteroduplexes could be a gene-conversion or a recombination event. These regions also have a high density of minisatellite loci, revealed by the fact that about 80% are found at the end of genetic linkage maps (NIH/CEPH, 1992) and *in situ* hybridisation localised most such loci to the terminal bands of chromosomes (Royle *et al.*, 1988). It is possible that locus specific minisatellites play a role in homologue recognition. The detailed human genetic maps of human chromosomes that have been generated in recent years, have revealed indirectly that the proterminal regions of human chromosomes have elevated recombination rates. Some of the markers used in linkage analysis have been placed on physical maps and comparison of the genetic and physical maps has revealed linkage map expansion at the ends of many chromosomes (NIH/CEPH, 1992; Litt *et al.*, 1995), the underlying cause of which must be elevated recombination frequencies. Male genetic maps are generally shorter than female maps except towards the ends where genetic map expansion is greater in males than in females (NIH/CEPH, 1992). In addition, chiasmata which represent recombination events are distributed preferentially towards the terminal regions of human chromosomes during diakinesis of meiotic prophase I (Hulten, 1974; Laurie and Hulten, 1985). The proterminal regions have a high density of CpG island clusters and genes (Saccone *et al.*, 1992;

Craig and Bickmore, 1994), which may be accounted for by the elevated recombination rates in proterminal regions. This is because double recombination and gene conversion events have the effect of generating homozygosity, which would be important in the removal of defective copies of genes. Therefore, it may be beneficial for genes to be located in the terminal regions of chromosomes (van Ommen, 1994).

The isolation of human subterminal sequences. Terminal sequences are generally not represented in conventional genomic libraries, but the majority of interstitial telomeric repeats are located in subterminal regions (Meyne *et al.*, 1989). Therefore, screening of cosmid libraries with a telomere repeat probe has resulted in the isolation of clones containing subterminal sequences (Wells *et al.*, 1990). Several specialised cloning strategies have been utilised to isolate subterminal and telomeric sequences from the human genome. These include a technique used by de Lange *et al.* (1990) which involved rendering the chromosome termini blunt ended with *Bal31* exonuclease digestion followed by cleavage with restriction enzymes which were unlikely to cut within the telomere repeat array. Size fractionation and blunt ended cloning into a plasmid vector followed by screening of the clones with a telomere repeat probe, yielded two isolates containing sequences adjacent to telomere repeats (de Lange *et al.*, 1990). Techniques based on the polymerase chain reaction have also been developed to isolate subterminal sequences in mice (Broccoli *et al.*, 1992). Another PCR-based technique for the isolation of sequences immediately adjacent to human telomeres involved the use of a primer designed to amplify from the G-rich strand of the telomere repeat array. This primer was used in conjunction with a primer designed to amplify from a linker ligated to *MboI* sites in genomic DNA digested with *MboI* and size selected for fragments larger than 5.5kb. Telomere-junction fragments were isolated by amplification followed by cloning into a plasmid vector and screening with a telomere repeat probe (Royle *et al.*, 1992). A similar technique but utilising a vectorette strategy was used to isolate sequences adjacent to telomeres in barley (Kilian and Kleinhofs, 1992). The most successful technique for isolating long segments of subterminal sequences involves the use of yeast artificial chromosome (YAC) vectors. The YACs used in this technique have been modified so they contain a single telomere. These 'half YACs' will only propagate fragments that contain telomere repeat units at the extreme end. This is because the presence of human telomere repeats can complement the function of a yeast telomere by acting as a template for the addition of yeast telomere repeats (Cross *et al.*, 1989; Brown, 1989; Riethman *et al.*, 1989; Cheng *et al.*, 1989).

Human subterminal sequences. Many different isolates of human subterminal sequences have been obtained and sequenced. These sequences contain several different repeats each of which exhibits a high degree of polymorphism. They vary according to sequence, number of tandem repeats and in their distribution within the genome (Brown *et al.*, 1990; de Lange *et al.*, 1990; Ijdo *et al.*, 1992). As for the subterminal sequences in yeast, the variable distribution suggests that these sequences in humans have no function. One major subterminal sequence family in humans is represented in several different subterminal clones (de Lange *et al.*, 1990; Wells *et al.*, 1990; Cross *et al.*, 1990; Brown *et al.*, 1990; Cheng *et al.*, 1991; Royle *et al.*, 1992). This sequence family contains three different types of repeated sequence. The most distal sequence is a discontinuous array of five 37bp repeats, proximal to this is a variable length 85% GC-rich array of a 29bp repeat unit and the most proximal array of tandem repeats is composed of a 61bp repeat which also varies in repeat number (Royle, 1995). The polymorphic distributions of these subterminal sequences could be explained by unequal exchange events between non-homologous chromosomes, such events could result in homogenisation of these sequences. However, substantial sequence divergence in the order of 2-22% exists within the sequence family (Cross *et al.*, 1990; Royle *et al.*, 1992). This suggests that homogenisation by unequal terminal exchanges between non-homologous chromosomes must occur at a low frequency or with a higher frequency between subsets of chromosomes. Some members of this sequence family are transcribed into a 6kb polyadenylated RNA, however, no long open reading frames have been seen in the probe used to detect the RNA transcripts so it is not known whether these transcripts encode proteins (Cheng *et al.*, 1991). Not all chromosome ends contain subterminal repeats, for example the subterminal region shared by Xp and Yp is unique in the human genome and other chromosome ends do not hybridise to human subterminal repeat sequence probes (Cross *et al.*, 1990; Brown *et al.*, 1990; de Lange *et al.*, 1990; Ijdo *et al.*, 1992). These chromosome ends may also be unique, they may contain highly diverged versions of the subterminal repeat family mentioned above, or they may contain an as yet uncharacterised subterminal sequence family. The presence or absence of subterminal sequences at the end of a chromosome can result in chromosome length polymorphisms. Length polymorphism has been detected at the end of the short arm of chromosome 16. At this locus four alleles exist which differ in length dramatically such that the distances of the telomere from the 5′ HVR are 170, 245, 350 and 430kb (Wilkie *et al.*, 1991; Harris and Thomas, 1992). The 170 and 430kb alleles hybridise to TelBam3.4 which contains a copy of the human subterminal sequence family described above and the 350kb allele hybridises to another subterminal sequence probe TelBam11 (Brown *et al.*, 1990). Distal to the point of divergence (approximately 145kb distal to the α-globin genes), the

sequences of the 16pter alleles have more similarity to other chromosome ends than they do to each other. Therefore, it is likely that these size polymorphisms have arisen as a result of unequal terminal exchanges between non-homologous chromosomes. The 16pter alleles are in strong linkage disequilibrium with alleles at polymorphic loci proximal to the point of divergence, which suggests that the exchange events that generate these polymorphisms must be rare and that the rate of recombination must be reduced in this region. Since the 16pter alleles are so different they presumably do not pair during synapsis and the linkage disequilibrium suggests that chromosome recognition must occur in a more proximal location (Wilkie *et al.*, 1991). Given the distribution of the subterminal repeat families in the human genome, terminal length polymorphisms are likely to be fairly common (Ijdo *et al.*, 1992) but to date they have only been demonstrated at 2q (Macina *et al.*, 1994) and at both ends of chromosome 16 (Harris and Thomas, 1992).

Subterminal sequences in other species. The human subterminal sequence family described above has been detected in the chimpanzee, gorilla and orang-utan genomes (Royle *et al.*, 1994; Hoglund *et al.*, 1995). A telomere-anchored PCR technique has been used to isolate subterminal sequences from the chimpanzee genome. One group of clones contained tandem arrays of a 32bp repeat unit, which showed extensive hybridisation to gorilla and chimpanzee DNA but was not present in the human and orang-utan genomes. *In situ* hybridisation revealed that this 32bp satellite sequence is present at the ends of most of the gorilla chromosomes and 21 of the 48 chimpanzee chromosome arms (Royle *et al.*, 1994). Sequences homologous to human subterminal sequences are present in these primate genomes, but must be present at a small number of chromosome ends or located proximal to the 32bp repeat satellite sequence. This information demonstrates that the ends of primate chromosomes must be dynamic structures which have undergone rapid and large scale changes over a short period of evolutionary time.

Despite the differences in the structure of the terminal sequences of *Drosophila* with those of most other eukaryotes, subterminal sequences that have similar properties have been isolated. The subterminal sequences of the *Drosophila* X chromosome are composed of tandem arrays of 1.8kb and 0.9kb repeats, these repeats are often truncated and contain internal deletions and subrepeats (Karpen and Spradling, 1992). *In situ* hybridisation demonstrates that the subterminal repeat from the *Drosophila* X chromosome cross hybridises to the terminal regions of chromosome 3R (Karpen and Spradling, 1992).

The subterminal regions of *Plasmodium falciparum* contain several repetitive sequences which are conserved between different chromosome ends (de Bruin *et al.*, 1994). These sequences are responsible for the large chromosomal size polymorphisms which range from 50 to over 300kb (Corcoran *et al.*, 1988). It has been suggested that these subterminal sequences are the preferential sites of chromosome pairing and promote recombination and gene conversion events. Such events would explain the conserved nature of these elements between chromosomes, and pairing between dispersed subterminal repeats coupled with exchanges of information could provide a means of rapidly generating phenotypic diversity (de Bruin *et al.*, 1994). Extreme genetic diversity is observed between different isolates of *Plasmodium falciparum* and is presumably necessary for this obligate intracellular parasite to evade the host's defence mechanisms.

Subterminal repeated sequences have been observed in a range of different organisms and generally speaking they are species specific. Although no specific function has been attributed to these regions within any one organism, they have been shown to promote genetic exchange which could have evolutionary consequences for the species in which they reside. For example, homologous exchanges in these regions may create homozygosity removing defective genes (van Ommen, 1994), or non-homologous exchanges of genetic information could promote genetic diversity (de Bruin *et al.*, 1994).

The function of telomeres

1. Complete replication of a linear DNA molecule. Semi-conservative replication of a linear DNA molecule poses a problem as to how the molecule can be completely replicated. When the replication fork reaches the end of the chromosome the leading strand can replicate to the end of the 5′ strand (Fig. 1.1.). The replicating lagging strand contains RNA primers that initiate the synthesis of Okazaki fragments however, as the RNA primer is removed from the lagging strand a single stranded gap is generated (Fig. 1.1.). This results in a truncation of the newly synthesised strand by a minimum length equivalent to the length of an RNA primer (8 to 14 nucleotides) and a maximum length equivalent to the length of an Okazaki fragment (150 to 200 nucleotides) (Thommes and Hubscher, 1990). Therefore, specialised systems are required to overcome this end replication problem.

a. Telomerase. Eukaryotic organisms have the ability to counterbalance the loss of terminal sequences by the *de novo* synthesis of sequences at the end of the chromosome. In those organisms that have the simple repeated sequence at the

Fig. 1.1. The problem of the complete replication of a blunt ended linear DNA molecule by semi-conservative replication

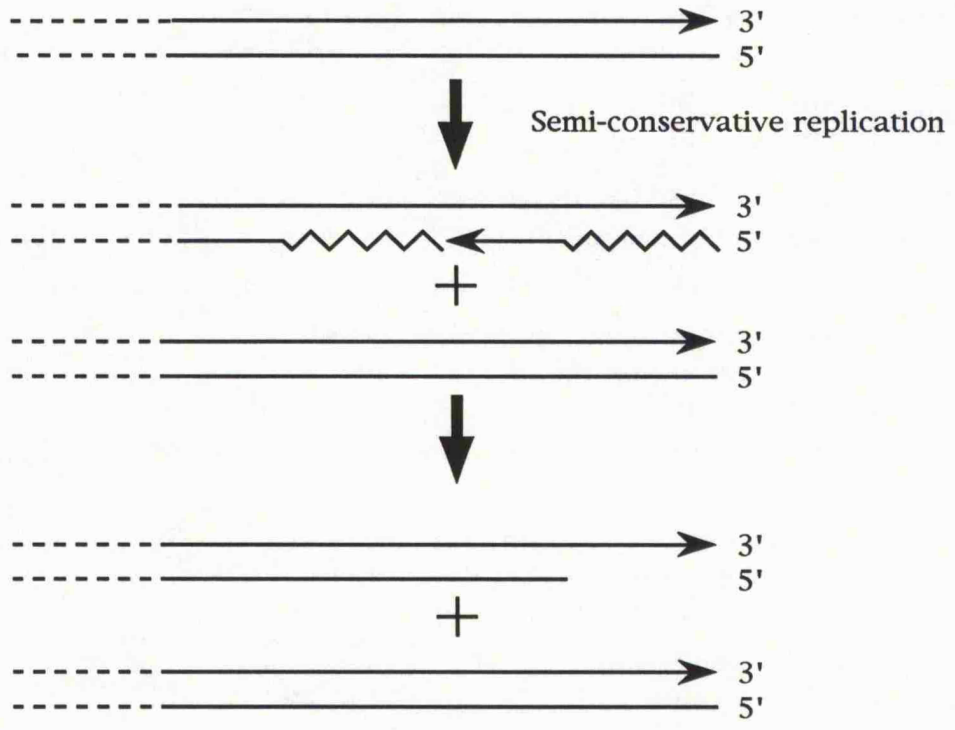
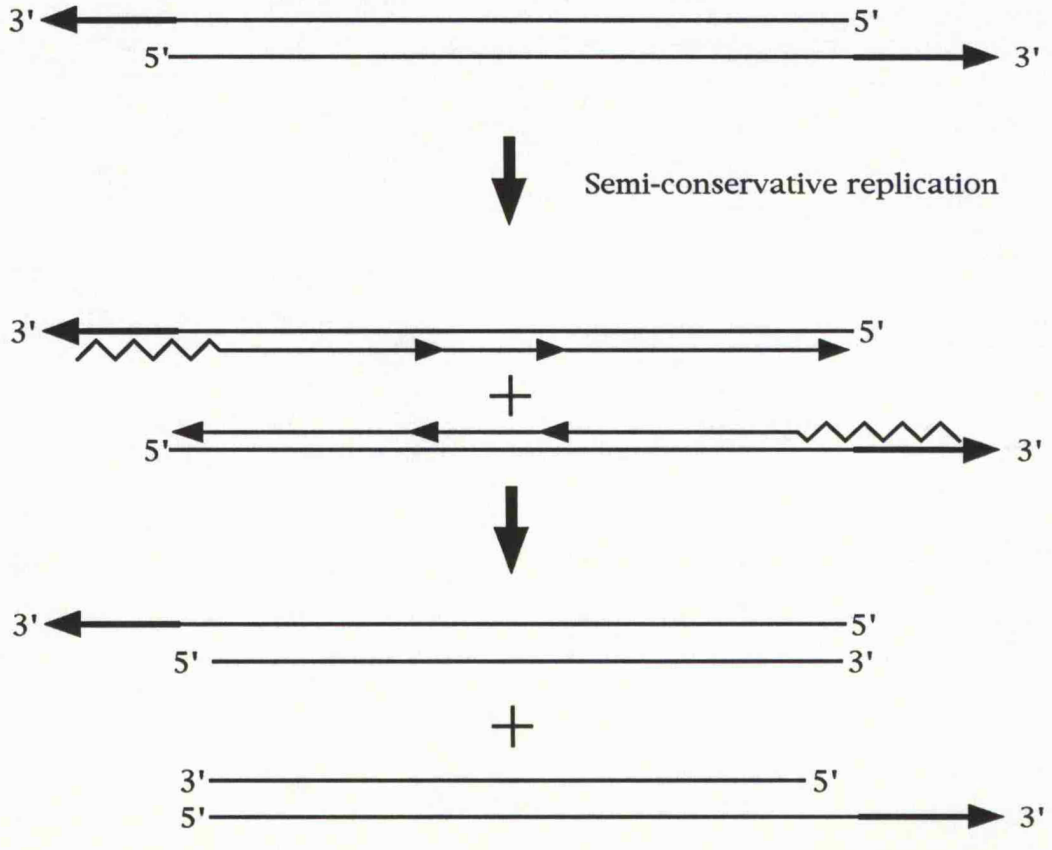


Fig. 1.2. The loss of the 3' terminal over hang during Semi-conservative replication



extreme end of the chromosome, this function is carried out by the enzyme telomerase. Telomerase is a ribonucleoprotein which catalyses the addition of the tandem telomeric repeat sequence to the 3′ end of the G-rich strand. Telomerase was first identified in *Tetrahymena* (Greider and Blackburn, 1985). Subsequently, telomerase activity has been detected in the ciliates *Oxytricha* and *Euplotes* (Shippen-Lentz and Blackburn, 1989), in yeast (Cohn and Blackburn, 1995), in mice (Prowse *et al.*, 1993) and in humans (Morin, 1989). Telomerase is a specialised reverse transcriptase containing an RNA component which serves as a template for the synthesis of telomere repeats. The RNA component has been identified in *Tetrahymena* (Greider and Blackburn, 1989), *Euplotes* (Shippen-Lentz and Blackburn, 1990), *Saccharomyces cerevisiae* (Singer and Gottschling, 1994), mice (Blasco *et al.*, 1995) and in humans (Feng *et al.*, 1995). The sequence of the RNA component of telomerase is not conserved between species. This is particularly striking when the sequence of the mouse and human RNA components are compared, they have just 65% homology with an 11 nucleotide template in humans and a 9 nucleotide template in mice, yet they have the same telomeric sequence (Blasco *et al.*, 1995; Feng *et al.*, 1995). The sequence divergence accounts for the difficulty in isolating the RNA component from higher eukaryotes. High divergence exists between the RNA components of different species of *Tetrahymena*, however, it has been proposed that although the primary structure is different they may in fact have a conserved secondary structure (Romero and Blackburn, 1991). If the RNA template is mutated the sequence of the telomere repeat unit in the mutant changes to that of the mutated template (Yu *et al.*, 1990). Two protein components of the *Tetrahymena* telomerase have been identified which lack extensive sequence homology to other polymerases and therefore appear to represent a novel type of polymerase (Collins *et al.*, 1995). There has been speculation that the *EST1* (Ever Shorter Telomeres) gene identified in *Saccharomyces cerevisiae* may code for a protein component of yeast telomerase (Cohn and Blackburn, 1995; Lin and Zakian, 1995). The sequence requirements of *Tetrahymena* telomerase to catalyse the addition of telomere repeats has been investigated. *Tetrahymena* telomerase does not require the presence of a sequence complementary to the RNA template sequence at the termini, in fact the presence of two or complementary repeats upto 36 bases from the 3′ terminus was sufficient for elongation (Harrington and Greider, 1991).

Heterogeneity is observed in the telomere repeat sequence of many organisms. The complete sequence of some *S. cerevisiae* telomeres have been obtained, this shows that the telomeric sequence varies TG_{1-3} (Shampay *et al.*, 1984) and that the variation extends to the terminus. The most likely explanation of this is that telomerase can 'stutter' during the synthesis of telomere repeats creating repeats

of irregular lengths (Cohn and Blackburn, 1995). The length of telomeres in some strains of mice are very long (Kipling and Cooke, 1990), however the telomerase activity in mice is nonprocessive, synthesising just a few repeats (Prowse *et al.*, 1993). This is in contrast to the situation in humans where the telomere repeat arrays are much shorter but the telomerase activity is more processive, synthesising many repeats (Prowse *et al.*, 1993). This suggests that factors other than just telomerase have a role in the regulation of telomere length. This discrepancy may also be accounted for by biochemical differences between mouse and human telomerase, where the sequences of the human and mouse telomerase RNA components show that the template region of the mouse RNA component (9 nucleotides) is shorter than that in the human component (11 nucleotides) (Blasco *et al.*, 1995; Feng *et al.*, 1995).

Although the role of telomerase is to compensate for the loss of terminal sequences incurred by semi-conservative replication, its role may actually be more specific than this. The termini of the chromosomes of many lower eukaryotes containing telomere simple repeats have been shown to have a 3′ terminal overhang (Klobutcher *et al.*, 1981; Henderson and Blackburn, 1989; Wellinger *et al.*, 1993; Vermeesch and Price, 1994) and it has been argued that the end replication problem lies in the complete replication of the leading strand to restore the 3′ terminal overhang (Fig. 1.2.) (Lingner *et al.*, 1995). The most likely mechanism for this is that telomerase catalyses the addition of the telomere repeats to the leading strand after semi-conservative replication. However, telomerase requires a single-stranded end as a substrate (Lee *et al.*, 1993) therefore in this model additional components may be required for telomerase to recognise and act upon a double-stranded end (Lingner *et al.*, 1995).

b. Recombination. The introduction of linear plasmids containing *Tetrahymena* telomere repeats at one terminus and *Oxytricha* repeats at the other in yeast reveals another mechanism for telomere maintenance. Sequences can be transferred from one end to the other without the loss of telomere repeats from the donor end, suggesting that gene-conversion like events mediate this exchange of information (Pluta and Zakian, 1989; Wang and Zakian, 1990). Yeast cells lacking *EST1* function exhibit a decline in the length of the telomere repeat array, chromosome loss and cellular senescence (Lundblad and Szostak, 1989). However, in late cultures of *est1* cells some cells overcome the apparently lethal effects of lacking *EST1* function. Some of these survivors result from the amplification of Y′ subterminal elements, the formation of such survivors is *RAD52* dependent (Lundblad and Blackburn, 1993). This implies that the acquisition of Y′ elements is recombination dependent and restores a functional telomere (Lundblad and Blackburn, 1993). Telomerase activity has been detected

in yeast and the RNA component has been identified (Shippen-Lentz and Blackburn, 1990; Cohn and Blackburn, 1995), therefore it appears that the primary mechanism for the maintenance of telomeres in yeast is by telomerase. However, the recombination pathways could be a secondary backup type of mechanism for the rescue of telomere function. It is also possible that similar recombination based pathways could exist in other organisms including humans given the presence of subterminal sequence families and elevated recombination rates in the proterminal regions.

c. **Transposition.** *Drosophila* chromosomes do not contain telomere repeat sequences and viable deletions of *Drosophila* chromosomes have been generated which loose terminal sequences at a rate of between 50-100bp per fly generation (Biessmann and Mason, 1988; Levis, 1989). This loss of terminal sequences is greater than that predicted to be the minimum loss of sequences through incomplete semi-conservative replication, between 34-38bp per fly generation (Levis, 1989). These deleted chromosomes contained no telomeric repeats (Levis, 1989). This means that these chromosome breaks are not stabilised by the addition of terminal sequences, such as that observed in other organisms which contain telomeric simple sequence. Therefore, in *Drosophila* the stabilisation by the capping of the broken chromosomes and the counterbalancing of the loss of terminal sequences by incomplete replication must be carried out by two separate mechanisms. Subsequently it was demonstrated, that the ends of natural *Drosophila* chromosomes contain retrotransposable elements termed HeT-A and TART (Biessmann *et al.*, 1992; Levis *et al.*, 1993). Incomplete DNA replication at the termini of *Drosophila* chromosomes is counterbalanced by transposition of these elements at a high frequency specifically to the termini. The length of DNA added by the 6kb HeT-A elements which transpose to 1% of chromosomes per generation is an average of 60bp per chromosome end per generation, this is adequate to counterbalance the loss of terminal sequences by incomplete replication (Biessmann *et al.*, 1992). These transposable elements are not responsible for capping chromosome ends in *Drosophila* to prevent cell cycle arrest and fusions, to date the structure performing this function is not known. It has been proposed that this capping function could be performed by a protein structure which specifically recognises double-stranded breaks in a sequence independent fashion (Biessmann and Mason, 1988).

2. **Chromosome capping.** It was clear from the work of Muller and McClintock that the terminus of a chromosome was 'capped' since broken chromosomes became unstable and tended to fuse to other broken ends (Muller, 1938; McClintock, 1938; Muller, 1940). The presence of a 'cap' on the end of the chromosome stabilised the chromosome, this function has been attributed to

telomeres. How telomeres carry out this function is unclear. Chromosome capping could be carried out by the 3′ overhang detected at the end of the telomere repeat arrays of many organisms. *In vitro* studies demonstrate that non-Watson-Crick base pairing between residues in the 3′ overhang of the G-rich strand can form an intramolecular foldback tetraplex or G-quartet structure (Williamson *et al.*, 1989; Hardin *et al.*, 1991; Wang *et al.*, 1994). But, the existence of G-quartet structures at the terminus of telomeric DNA has not been demonstrated *in vivo*. However, it may be more than just a coincidence that all telomeric repeat sequences are based upon GGGG or AGGG which are the sequences that can best adopt the G-quartet structure (Murchie and Lilley, 1994). If these structures are present *in vivo* then the evolutionary conservation of the telomeric simple sequence could be accounted for by a functional constraint operating to maintain the sequences that can best adopt the G-quartet structure. In the somatic tissues of some organisms, for example humans, the enzyme telomerase is not active. This means that semi-conservative replication of the terminus will result in the lagging strand having a 3′ overhang and the leading strand would be blunt ended (Fig. 1.2.). As a result, only half of the telomeres of the somatic tissues of some organisms have a 3′ terminal overhang. The role of the 3′ overhang in the capping of the terminus may not be significant in tissues where telomerase is not active. The *Oxytricha* telomerase is not able to use $(TTTTGGGG)_n$ as a substrate for the synthesis of telomere repeats when this oligonucleotide is in an intramolecular foldback structure (Zahler *et al.*, 1991). Also the inhibition of intramolecular foldback structures by the use of base analogues which cannot form G-G base pairs does not prevent the *in vivo* healing by telomerase in yeast (Lustig, 1992). Proteins have been identified which recognise the structure of G-quartet structures (Weisman-Shomer and Fry, 1993), these bound proteins could in themselves provide the capping function or could mediate the interaction of the terminus with the nuclear architecture (Gottschling and Cech, 1984; de Lange, 1992). The two subunit telomere binding protein identified in *Oxytricha* binds as a heterodimer to the single-stranded 3′ overhang protecting it from chemical modification and exonucleotic degradation (Gottschling and Zakian, 1986). A single subunit protein with a similar binding activity has been isolated from *Euplotes* (Price, 1990). Both of these telomere binding proteins bind the 3′ overhang with a high affinity, such that in the salt concentrations in which all other DNA binding proteins disassociate these proteins remain non-covalently attached (Price *et al.*, 1992). This protein cannot bind oligonucleotides which are folded into an intramolecular foldback structure (Raghuraman and Cech, 1990). The kinetic inaccessibility of G-quartet structures suggests that additional factors are required for their formation *in vivo* (Pilch *et al.*, 1995). Subsequently, it was discovered that the beta subunit of the *Oxytricha* telomere binding protein promotes the formation of G-quartet structures (Fang

and Cech, 1993). Also the RAP1 telomere associated protein isolated from yeast can promote the formation of G-quartet structures and can form a stable complex with preformed G-quartets (Giraldo *et al.*, 1994; Giraldo and Rhodes, 1994). It is clear that proteins are necessary in mediating the capping function of telomeres, however the role, if any, of G-quartet structures *in vivo* is unclear. Other proteins have been identified that are important in telomere function particularly in the formation of specific telomeric chromatin which may have a role in the protection of telomeres from degradation, these proteins are discussed later.

The loss of telomeric sequences, and hence the capping function, results in an increase in chromosomal aberrations such as dicentric chromosomes as observed in senescing cell cultures (Benn, 1976; Sherwood *et al.*, 1988) and in tumours (Fitzgerald and Morris, 1984; Pathak *et al.*, 1988). Chromosomal fusions could have consequences for genome evolution, for example, the human chromosome 2 is thought to be a relic of a telomere-telomere fusion between two ancestral chromosomes, the fusion point has been cloned and consists of a head to head array of telomere repeats (Ijdo *et al.*, 1991).

3. Nuclear organisation. Human telomeres are tightly associated with the nuclear matrix during mitotic interphase (de Lange, 1992) and similar associations have been observed in some plants (Rawlins *et al.*, 1991). Telomeres in mice and trypanosomes appear to be distributed throughout the interphase nuclei (Chung *et al.*, 1990; Vourc'h *et al.*, 1993). In contrast to mitotic nuclei, telomere associations with the nuclear envelope are common in meiotic nuclei and result in a chromosomal formation called the bouquet structure. This formation may promote homologue recognition (reviewed by Loidl, 1979). In vegetative cells of the fission yeast *Schizosaccharomyces pombe* the centromeres are associated with the spindle pole body (fungal equivalent of the centrosome) prior to meiosis (Funabiki *et al.*, 1993). However, just prior to meiosis the telomeres associate with the spindle pole body and lead chromosomal movement until chromosome segregation starts. Then the centromeres resume the leading position in chromosome movement (Chikashige *et al.*, 1994). The clustering of the chromosomes is similar to the bouquet structure, suggesting that telomere-led premeiotic chromosomal movements may be a general feature in eukaryotes (Chikashige *et al.*, 1994).

4. Healing of broken chromosomes. Telomerase activity is responsible for the healing of broken chromosome ends such as that observed during macronuclear development in *Tetrahymena* where the chromosomes fragment and the ends are healed by the *de novo* addition of telomere repeats by telomerase

(Yu and Blackburn, 1991). A similar developmentally controlled healing, probably by telomerase, occurs in the early *Ascaris* embryo (Muller *et al.*, 1991). The *de novo* healing by telomerase of broken chromosomes is thought to be responsible for the stabilisation of truncated human 16p chromosomes by addition of TTAGGG repeats *de novo* (Wilkie *et al.*, 1990; Lamb *et al.*, 1993; Flint *et al.*, 1994). One of these 16p truncation sites (Wilkie *et al.*, 1990) was subsequently shown to be a recognition site for telomerase which catalysed the addition of TTAGGG telomere repeats *in vitro* (Morin, 1991). The fact that telomerase can heal broken chromosome ends and does not just catalyse the addition of telomere repeats to an established telomere, suggests that the sequence specificity for telomerase binding may not be very high. The comparison of the sequences of breakpoints that have been healed by telomerase reveals some similarity to the RNA template of telomerase (Flint *et al.*, 1994). This is concordant with *in vitro* studies on the sequence requirement of the *Tetrahymena* telomerase (Harrington and Greider, 1991). It has been proposed that exonucleases may degrade double stranded ends until a suitable template for telomerase is revealed (Kramer and Haber, 1993). If such exonuclease activity occurred in humans, only a few bases would be lost since the putative sequence requirement would occur often (Flint *et al.*, 1994). It is also possible that chromosome healing in humans could be mediated by recombination such as that observed in yeast (Pluta and Zakian, 1989; Lundblad and Blackburn, 1993). A possible example of this was described by Flint *et al.* (1994) where the break point of a deleted chromosome had additional sequences not present in the original DNA or the telomere sequence, suggesting this could be derived from another chromosome by recombination. In addition, this telomere contained a variant repeat (TTTAGGG) which would not be expected to occur in a telomere recently synthesised by telomerase.

The fact that human proterminal regions are rich in genes (Saccone *et al.*, 1992; Craig and Bickmore, 1994) and that telomerase activity or recombination is sufficient to stabilise broken chromosome ends, suggests that a significant proportion of human genetic diseases could arise as chromosome deletions. To give rise to viable offspring such deletions would have to be fairly small so that only a few genes were lost. Chromosomal abnormalities contribute to 40% of cases of severe mental retardation and 10-20% of mild mental retardation, however the resolution of the cytogenetic techniques routinely used to detect these abnormalities means that rearrangements of less that 1-2Mb may not be detected (Flint *et al.*, 1995). It is therefore possible that a large proportion of mental retardation is a result of cryptic chromosomal rearrangements and will only be detected by molecular genetical techniques. A scan for cryptic subterminal rearrangements using molecular genetic markers at the ends of

chromosomes reveals that such rearrangements account for at least 6% of unexplained mental retardation (Flint *et al.*, 1995).

The ability of telomerase to heal chromosome ends has been utilised to generate deleted chromosomes for mapping purposes (Farr *et al.*, 1992). This is done by the introduction of linear constructs containing telomeric sequences into somatic cell hybrids containing human chromosomes with selection for the loss of a distal marker (Farr *et al.*, 1992). Chromosome fragmentation occurs at the sites of integration of the construct and the chromosome is stabilised by the extension of the telomere repeats. It is also possible to target chromosomal breakage to a specific region through the use of gene targeting (Itzhaki *et al.*, 1992).

The healing of broken chromosome ends by telomerase could have implications in the genome evolution, allowing the creation of new chromosomal termini. This may have occurred at human telomeres when the subterminal satellite which may have existed in the progenitor to the human, chimpanzee and gorilla lineages (Royle *et al.*, 1994) was lost from the human lineage and new telomeres seeded. This raises the possibility that the majority of human telomeres have been formed fairly recently in the context of evolutionary time.

Telomere associated proteins. Proteins have been identified in various organisms which are essential for telomere function. The best characterised telomeric proteins are in the lower eukaryotes such as the hypotrichous ciliates and yeast. This is primarily because these organisms are more amenable to genetic analysis. As mentioned earlier, telomere binding proteins have been identified in Euplotes and Oxytricha, these bind with a high affinity to the 3′ single-stranded overhang and protect the chromosome end from exonucleotic degradation (Gottschling and Zakian, 1986; Price, 1990; Price *et al.*, 1992).

In Yeast. Telomeric DNA in *Saccharomyces cerevisiae* is organised into a non-nucleosomal chromatin structure called the telosome, in contrast to the subterminal repeats X and Y' which are assembled in nucleosomes (Wright *et al.*, 1992). The telosome does not appear to contain protein with as high an affinity as those isolated from ciliates (Wright and Zakian, 1995). The major protein component of the telosome is RAP1 (Wright *et al.*, 1992). RAP1 (repressor/activator protein) is an abundant DNA binding protein involved in control of transcription of several genes. RAP1 plays a part in the control of the expression of the silent mating type loci (Buchman *et al.*, 1988). The binding of RAP1 to the *HIS4* gene is required for the high levels of meiotic recombination observed at this locus (White *et al.*, 1991). This may be achieved by the ability of RAP1 to untwist double-stranded DNA (Gilson *et al.*, 1994). RAP1 is also a telomere

binding protein, binding the telomere repeat array approximately every 18bp (Gilson *et al.*, 1993; Graham and Chambers, 1994). The binding of RAP1 to telomeric DNA renders it resistant to DNase I and introduces 90° bends which could have an effect on the chromatin structure (Gilson *et al.*, 1993). Temperature sensitive mutations in the *RAP1* gene demonstrate its role in the maintenance of telomeres. When these mutants were grown at the semi-permissive temperature the length of the telomere arrays shortened and then stabilised (Conrad *et al.*, 1990). RAP1 is also a component of the nuclear scaffold in yeast (Cardenas *et al.*, 1990), this is relevant as the telomeres in humans are tightly associated with the nuclear matrix in all stages of interphase of several cell types (de Lange, 1992).

Other proteins have been identified in yeast which interact with yeast telomeres. The expression of the C-terminus of RAP1 in a strain with normal RAP1 activity results in changes in telomere length and chromosome stability (Conrad *et al.*, 1990). This suggested that some component was interacting with RAP1, which was subsequently identified to be RIF1 (RAP1 interacting factor) (Hardy *et al.*, 1992). Strains carrying disruptions of the *RIF1* gene are defective in transcriptional silencing and telomere length regulation but grow normally (Hardy *et al.*, 1992). It has not been determined whether RIF1 binds DNA directly or is a part of the telomeric heterochromatin via interactions with RAP1. There are also *trans* acting protein components, which are involved in telomere maintenance, the most important of which is thought to be EST1. Cells carrying mutations in the *EST1* gene progressively lose telomeric DNA with each cell division and eventually undergo cellular senescence (Lundblad and Szostak, 1989). As mentioned earlier, some cells with an *est1* genotype can survive senescence due to a recombination based pathway which rescues telomeric function (Lundblad and Blackburn, 1993). This phenotype is concordant with what one would anticipate for an essential component of the yeast telomerase, and indeed the *EST1* gene is the best candidate identified so far for a protein component of the yeast telomerase. Telomerase activity has been detected in yeast but is absent in *est1* cells suggesting that *EST1* is an essential component of telomerase (Lin and Zakian, 1995). Conversely, another study revealed that *EST1* activity was not required for telomerase activity (Cohn and Blackburn, 1995). Although *EST1* is an essential component of the yeast telomere maintenance mechanisms its precise role has not been identified.

In other species. Proteins have been identified which bind to telomeric sequences in mammals, however, with the exception of the RNA components of telomerase (Blasco *et al.*, 1995; Feng *et al.*, 1995) few genes have been identified which affect telomere length maintenance. An abundant protein (A2/B1) has

been identified in mice which binds to single stranded telomeric DNA (Mckay and Cooke, 1992). However, a telomeric location for this was not observed in metaphase chromosomes (Mckay and Cooke, 1992). It has been suggested that this protein may be involved in the processing of primary RNA transcripts (Ishikawa *et al.*, 1993). It therefore seems likely that A2/B1 is not involved in telomere maintenance. Telomere binding activity has been observed in mammals by a 50kDa protein called the TTAGGG repeat factor (TRF) (Zhong *et al.*, 1992). TRF only binds to double stranded oligonucleotides of six or more TTAGGG repeats and will not bind to related repeats such as TTTAGGG and TTGGGG (Zhong *et al.*, 1992). Immunofluorescence microscopy revealed that in interphase and metaphase nuclei the TRF protein occupies the chromosome ends (Chong et al., 1995). The TRF protein contains Myb-type helix-turn-helix DNA binding motifs which are also present in the RAP1 protein of yeast (Chong et al., 1995). Therefore, the TRF protein appears to be a telomere binding protein specific to the mammalian telomere repeat TTAGGG, the contribution of this protein to telomere function has yet to be determined. Given the conserved nature of telomeric simple sequence, it was hoped that the proteins identified in the function of the telomeres in lower eukaryotes would also be conserved in higher eukaryotes, and so facilitate the identification of similar proteins. However, unlike the telomeres of lower eukaryotes, the telomeres of mammals appear to have a nucleosomal chromatin structure (Makarov *et al.*, 1993). In humans, telomeres may have a bipartite chromatin structure with unusual chromatin near the telomere terminus and a more canonical nucleosomal organisation in the proximal part of the telomere (Tommerup *et al.*, 1994). There is evidence to suggest that genes involved in telomere maintenance are conserved between humans and yeast. Cultured cells from individuals with ataxia telangiectasia (AT) have increased levels of chromosome instability with the common chromosomal aberration being end-to-end fusions (Kojis *et al.*, 1991). It is possible that AT cells many have some defect in telomere maintenance. The gene responsible for AT has been cloned (Savitsky *et al.*, 1995) and has homology to the yeast *TEL1* gene (Greenwell *et al.*, 1995). Mutations in the *TEL1* gene result in shortened telomeres. It is therefore possible that the AT gene could be responsible for telomere maintenance in humans, however, the functional relationships between these two genes have yet to be determined.

Many genes have been identified which if mutated affect the length of telomeres, though how the length of a telomere is controlled is unclear. It must be more complex than a simple equilibrium between loss of telomeric sequences through semi-conservative replication and addition by telomerase. Such a mechanism would have to be very precisely controlled since any deviation would result in a gradual change in length. There must be a system that can measure changes in

telomere length and adjust the maintenance activity accordingly. One system may be to measure the length of a telomere by the titration of a telomere binding protein. This is probably unlikely, since the introduction of a large number of additional telomeric sequences in yeast results in a small increase in the length of the telomere repeat array (Runge and Zakian, 1989). Also, such a system would only monitor the length of all the telomeres within a cell and therefore could not monitor the change in length of a single telomere. The length of *Euplotes* macronuclear telomeres is very tightly regulated at 42 nucleotides; the most internal 28 nucleotides is double stranded and the terminal 14 nucleotides form the 3′ overhang (Klobutcher *et al.*, 1981). However, newly synthesised macronuclear telomeres are longer and heterogeneous in length and they are trimmed by an active processing event (Vermeesch *et al.*, 1993). It is possible that a nuclease carries out the trimming, but exactly how the length is so precisely controlled is unknown. A single stranded binding protein such as the *Euplotes* telomere binding protein could provide protection from the nuclease (Vermeesch and Price, 1994). The mRNA for the *Euplotes* telomere binding protein accumulates at the time of telomere processing (Price *et al.*, 1994). Telomere length in humans is heterogeneous, it is inversely correlated with age and there is heterogeneity in telomere length between individuals of the same age but telomere length is highly heritable (heritability= 78%) (Slagboom *et al.*, 1994), suggesting that there is a strong genetic component in the maintenance of telomere length. To date the only human gene known to affect telomere length is the AT gene (Savitsky *et al.*, 1995).

Telomere position effect. The expression of genes can be affected by their chromosomal position and the relocation of a gene from euchromatin to heterochromatin can result in transcriptional silencing (reviewed in Kornberg and Lorch, 1992). This position dependent gene expression is known as position effect variegation (reviewed in Henikoff, 1990). In yeast, the presence of genes in a position adjacent to a telomere can result in transcriptional silencing, this phenomena is known as the telomere position effect. The extent of silencing depends upon the strength of the promoter, the distance from the telomere and the length of the telomere (Gottschling *et al.*, 1990; Renauld *et al.*, 1993; Kyrion *et al.*, 1993). The yeast telomere position effect also involves a number of proteins including SIR2, SIR3 and SIR4 as mutations in the genes encoding these proteins reverses the transcriptional repression of genes adjacent to a telomere (Aparicio *et al.*, 1991). The genes involved in telomere position effect also affect the transcriptional silencing observed at the silent mating-type loci involved in the determination of yeast mating type (Aparicio *et al.*, 1991). Transcriptional silencing may be a result of the formation of an altered chromatin structure at the silent mating-type loci and at telomeres (Park and Szostak, 1990; Gottschling,

1992). The role if any, of the telomere position effect in control of gene expression is unknown. Genes have been found near natural yeast telomeres, however, there is no data available to suggest that their position near telomeres affects their expression. It is most likely that telomere position effect is a biological by-product of a specialised protein-DNA complex (the telosome) on the accessibility of proteins involved in transcription. To date, telomere position effect on gene expression has not been observed in humans (Bayne *et al.*, 1994).

Telomere dynamics.

1. *In vitro*. Cell lines derived from humans can only survive for a limited number of cell divisions, this is known as the replicative potential (reviewed by Goldstein, 1990). After this limited number of cell divisions the cells cease to divide, this point has been termed mortality stage 1 (M_1) and is thought to be a cell cycle check point. M_1 can be overcome and the life span of these cultures extended by transformation with certain viral proteins. Although an extended life span can be achieved by transformation, eventually mortality stage 2 (M_2) is reached when the cell population stabilises such that cell death equals cell division. Ultimately the cells cease to divide and at this point the cells are said to be in crisis. At the crisis point the culture starts to die, but at a very low frequency cells can overcome crisis and become immortal (DiPaolo, 1983; Chang, 1985). Fusion of senescent cells with immortal cells results in the senescent phenotype, suggesting that senescence is the result of a dominant *trans*-acting factor (reviewed by Norwood *et al.*, 1990). These observations have been used as a model for cellular senescence *in vivo* and may have particular relevance in tissues where the cells have a high rate of turnover, for example, blood and skin cells.

The analysis of telomere length in primary fibroblast cultures reveals a progressive shortening at a rate of approximately 48bp per population doubling (Harley *et al.*, 1990; Levy *et al.*, 1992). Analysis of telomere loss in human embryonic kidney (HEK) cells cultured *in vitro* revealed a loss of approximately 65bp per population doubling (Counter *et al.*, 1992), and in culture peripheral blood lymphocytes the telomere loss was approximately 133bp per population doubling (Vaziri *et al.*, 1993). There is variation in the replicative capacity of human primary cells, and it was shown that this was proportional to telomere length of the donor, and was estimated to be 10 population doublings per kilobase of telomeric sequence (Allsopp *et al.*, 1992; Allsopp and Harley, 1995). The loss of telomeric sequences in these somatic cell cultures is probably due to incomplete replication. The average length of the telomeres in HEK cells at crisis was 1.5kb

and it seems likely that some of these cells may contain telomeres with little or no telomeric sequences. The loss of telomeric sequences is also correlated with an increase in the frequency of dicentric chromosomes, presumably caused by an absence of a telomeric cap preventing end-to-end fusions (Counter *et al.*, 1992). Such chromosomal fusions have been observed in senescent cells (Benn, 1976; Sherwood *et al.*, 1988). In untransformed and transformed HEK cells telomerase activity was not detected, however, in cell cultures which had become immortal, telomerase activity was detected (Counter *et al.*, 1992). Telomerase activity has been detected in a variety of different immortal human cell lines and so leads to the hypothesis that telomerase activation is required for immortality (Counter *et al.*, 1994). However, immortal cell lines have been observed which do not contain detectable telomerase activity. All of these had long telomeres of upto 50kb (Kim *et al.*, 1994; Bryan *et al.*, 1995), suggesting that an additional mechanism may be present for the stabilisation of telomeric sequences, for example, by a recombination based system involving subterminal sequences such as that observed in yeast (Lundblad and Blackburn, 1993). The analysis of the dynamics of a single telomere repeat array in a human immortal cell line without detectable telomerase activity, revealed a gradual shortening at a similar rate to that detected in mortal somatic cell lines (Murnane *et al.*, 1994). The telomeres were generally not lost completely and when the length of the array was a few hundred base pairs in length there was a sudden increase in length. In addition to a gradual shortening of telomere length large decreases were observed, it was proposed that these large increases or decreases in length were due to non-reciprocal recombination restoring telomere length (Murnane *et al.*, 1994).

2. ***In vivo.*** In humans the length of telomeres in somatic tissues is on average less than that of germline and foetal tissues (Cooke and Smith, 1986; Hastie *et al.*, 1990; de Lange *et al.*, 1990). In lymphocytes, the reduction in telomere length was estimated to be 33bp per year (Hastie *et al.*, 1990) and in skin the rate of telomere loss was estimated at 15bp per year (Allsopp *et al.*, 1992). Therefore, in human somatic tissues the telomere length is inversely correlated with age. Telomere length in trisomy 21 individuals is reduced compared to age-matched normal individuals (Vaziri *et al.*, 1993). It has been suggested that the loss of telomeric sequences may play a role in the premature immunosenescence observed in trisomy 21 patients. In addition, a reduction in telomere length and *in vitro* replicative capacity has been detected in individuals with the Hutchinson-Gilford syndrome of premature ageing (Allsopp *et al.*, 1992). The length of telomeres appears to be a good marker for the number of cell divisions it has taken to form a particular tissue. In addition to the age related loss of the telomeric sequences in somatic tissues, there is also an increase in the frequency of dicentric chromosomes (Bender *et al.*, 1989). These observations have led to the suggestion

that telomerase is inactive in somatic tissues but active in the germline (Hastie *et al.*, 1990). In support of this a small increase in the length of telomeres in the male germline has been detected (Allshire *et al.*, 1988). However, although telomerase activity has not been detected in many human somatic tissues (Kim *et al.*, 1994), low levels of activity have been detected in leukocytes (Broccoli *et al.*, 1995; Counter *et al.*, 1995; Hiyama *et al.*, 1995). In contrast to humans, high levels of telomerase activity has been detected in many mice tissues (Prowse and Greider, 1995), and primary cells lines derived from mice spontaneously immortalise without viral transformation (Macieira-Coelho and Azzarone, 1988).

3. **Tumours.** The analysis of telomeres in human tumour tissues reveals similarities to telomeres in cultured cells *in vitro*. It has been observed that telomeres are often shorter in tumour tissues when compared to normal tissues, for example, telomeres in colorectal carcinomas are usually shorter than in the normal colonic mucosa. Several tumours had very short telomeres ranging from 3kb to 1kb or less with some telomeres that may have completely lost telomere repeats (Hastie *et al.*, 1990). Telomere shortening has been detected in a large number of tumours including childhood leukaemia (Adamson *et al.*, 1992), giant cell tumour of bone (Schwartz *et al.*, 1993) and renal cell carcinoma (Mehle *et al.*, 1994). However, some tumours showed no reduction in telomere length and a few showed an increase (Schmitt *et al.*, 1994; Wainwright *et al.*, 1995). Also, telomeric associations have been observed in many tumours, which could be accounted for by the lost of normal telomere function in these tissues (Fitzgerald and Morris, 1984; Pathak *et al.*, 1988). It appears that like normal somatic tissues the early stages of many tumours lack an efficient mechanism of telomere maintenance. Therefore, by analogy with the immortalisation of culture cells *in vitro*, the activation of telomerase may be required for the progression of a tumour. Telomerase activity has been detected in a number of tumour tissues and its activation appears to be correlated with the acquisition of malignancy (Kim *et al.*, 1994; Counter *et al.*, 1994, 1995; Chadeneau *et al.*, 1995). Speculation about the use of telomerase as a target for chemotherapeutic agents in the treatment of cancer has followed these observations, and such speculation has been facilitated by the recent characterisation of the RNA component of the human telomerase (Feng *et al.*, 1995). The use of antisense RNA against the RNA component of human telomerase in an immortal cell culture resulted in telomere shortening and senescence after 23 to 26 population doublings (Feng *et al.*, 1995). This suggests that telomerase may indeed be a useful chemotherapeutic agent for cancer. However, such an agent may be effective only if a large proportion of the tumour cells are expressing telomerase and even if that were the case there may be a delay before a significant proportion of tumour cells are killed (Kipling, 1995). Several cell doublings would have to occur before sufficient terminal sequences

were lost resulting in cell death, consequently the tumour may become a life threatening size by the time a telomerase inhibitor took effect. Also, the telomerase inhibitor would have to be targeted specifically to the tumour, since in males the inhibition of telomerase in the germline would result in shortened telomeres in sperm cells, which may have consequences for the next generation. In addition, the inhibition of telomerase activity could affect the replicative capacity of the somatic cells in which telomerase is expressed. Telomerase activity has been detected in peripheral blood and bone marrow leukocytes (Counter *et al.*, 1995; Broccoli *et al.*, 1995), limiting the replicative capacity of these cells could perhaps result in premature immunosenescence.

The telomere hypothesis of cellular ageing. These data from humans on telomere dynamics both *in vitro* and *in vivo* have led to the telomere hypothesis of cellular ageing (Harley, 1991). In this hypothesis the M_1 checkpoint in cell growth is triggered by one or more telomeres reaching a critical length. The telomere shortening, perhaps accompanied by a double stranded break or the deletion of a gene or genes in the subterminal regions, could trigger an M_1 cell cycle arrest. The loss of subterminal sequences has been observed during the culturing of lymphocytes (Guerrini *et al.*, 1993). It has also been suggested that the shortening of telomeres could result in the repression of telomere adjacent genes, which in turn could trigger cell cycle arrest (Wright and Shay, 1992). Although the telomere position effect has been observed in yeast (reviewed in Sandell and Zakian, 1992) no such position effect has been observed in humans (Bayne *et al.*, 1994) where the bulk of telomeric chromatin is of a more canonical nature (Tommerup *et al.*, 1994). In certain circumstances the M_1 checkpoint can be overcome, for example, by viral oncogenes. In addition, breakage/fusion cycles initiated by telomeric associations have been implicated in the deletions on chromosome 11 resulting in a loss of heterozygosity during the progression to Wilms' tumour (Sawyer *et al.*, 1994). The bypassing of the M_1 checkpoint can compromise chromosome stability and telomeres continue to loose sequences until the M_2 checkpoint. At M_2 there is an increase in chromosome instability and in cell death but this can only be overcome by the activation of telomerase and the subsequent stabilisation of the chromosomes. The telomere hypothesis of cellular ageing may not play a major role in mice, since the telomeres of some mouse strains are long (Kipling and Cooke, 1990) and high levels of telomerase activity have been detected in somatic tissues (Prowse and Greider, 1995). The differences between humans and mice in the regulation of telomerase activity in somatic tissues, may be a consequence of the difference in longevity between the two species. Long-lived species such as humans may have additional mechanisms that control cellular proliferation and one of these may be the repression of telomerase in somatic tissues (Prowse and Greider, 1995). The M_1 and M_2

checkpoints regulate cellular proliferation and as a result may play a role in the protection against cancer.

Yeast cells also age, the number of cell divisions varies between strains. This was originally thought to be due to the accumulation of bud scars on the surface of the mother cells. However, recent evidence reveals that genetic factors are involved in the ageing process in yeast. A screen for mutations that affected life span revealed a number that had an elongated life span and four genes were identified called *UTH1-4* (Kennedy *et al.*, 1995). Characterisation of the *UTH2* gene revealed that it was the previously identified *SIR4* gene involved in transcriptional repression at HM loci and telomeres. It was suggested that certain mutant alleles of *SIR4* prevented the accumulation of SIR4 protein at the HM loci and telomeres, resulting in the accumulation of SIR4 in other regions of the genome such as the proposed AGE locus. It was proposed that silencing at the AGE locus by SIR4 and other proteins breaks down in old cells resulting in senescence (Kennedy *et al.*, 1995). The average telomere length in cultures carrying the elongated life span *SIR4* alleles is slightly reduced but is also reduced in cultures carrying other *SIR4* mutant alleles. Ageing in yeast is not a function of telomere loss but coincidentally involves proteins that interact with telomeres and the HM loci.

Human telomeres contain variant repeats

In addition to TTAGGG human telomeres contain different types of telomere repeats such as TGAGGG, TTGGGG and TTTAGGG as shown by hybridisation (Richards and Ausubel, 1988; Allshire *et al.*, 1988, 1989) and by sequence analysis of cloned telomeres (Brown *et al.*, 1990) (N. J. Royle pers. comm.). Hybridisation analysis revealed that TGAGGG repeats occurred within the most proximal 1.9kb and that the majority of TTGGGG repeats occurred within the most proximal 2kb of human telomere repeat arrays (Allshire *et al.*, 1989). The bulk of the remainder of the telomere repeat arrays consists of the human telomeric sequence TTAGGG. The significance of these telomere variant repeats and their distribution in the proximal regions of human telomeres is unknown. Telomere variant repeats may be functionally neutral and able to spread within and between telomeres by mechanisms involved in repeat turnover. The proximal location could be due to the loss of variant repeats by truncations in telomeres coupled with the subsequent healing by telomerase, resulting in the replacement of variant repeats with TTAGGG repeats. It was speculated that variant repeats could be involved in telomere function perhaps by the interaction with telomere proteins (Allshire *et al.*, 1989). There is evidence to suggest that there may be functional constraints preventing the spread of variant repeats throughout telomere repeat

arrays. The presence of variant repeats in proximal regions of *Tetrahymena* telomeres is tolerated with little change in phenotype (Yu and Blackburn, 1991). However, the presence of a completely altered telomeric sequence results in changes in telomere length, impaired cell division and cellular senescence (Yu *et al.*, 1990). In addition, the mammalian telomere binding protein TRF does not bind well to telomere repeats other than TTAGGG (Zhong et al., 1992). Also, DNA constructs containing TTAGGG repeats readily form telomeres upon transfection into human cells, however TTAGGG related repeats formed telomeres with a considerably lower frequency (Hanish *et al.*, 1994). Therefore, it is possible that functional constraints are acting on the variant repeats in humans preventing their spread throughout the telomere repeat array. It is also interesting to note that the average length of human telomeres in cultured primary cell lines at crisis is approximately 1.5kb (Counter *et al.*, 1992) which is approximately the distal limit of variant repeats in human telomeres (Allshire *et al.*, 1989). It may be possible that the telomere functions are not supported by telomere variant repeats, so that if the length is reduced such that the terminus is within the region covered by variant repeats telomere function is lost.

Other repeated DNA sequences in the human genome

The size of the haploid genome, known as the C-value, increases with the complexity of the organism in which it is measured. For example, in prokaryotes C-values of approximately 10^6bp have been observed and in higher eukaryotes C-values increase to 10^8-10^{11}bp. However, in higher eukaryotes C-values do not always correlate with the complexity of the organism, for example, the genome size of amphibians is approximately 25 times the size of those of mammals (Mirsky and Ris, 1951). This is known as the C-value paradox. Estimates of the number of genes in eukaryotic organisms revealed that there was a large discrepancy between the number of genes and genome size, suggesting that large regions of eukaryotic genomes consisted of additional DNA not related to gene function. Analysis of the reassociation kinetics of shared DNA fragments revealed that a large proportion of eukaryotic genomes consisted of sequences present in more than one copy (Britten and Kohne, 1968). Therefore, the C-value paradox arises from the variation in the quantity of non-coding DNA, large of proportions of which is repeated.

Dispersed repeats. There are two main classes of dispersed repeats comprising some 5% of the human genome, long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) (Singer, 1982). The most common LINE sequence in the human genome is the L1 element (6.4kb). The most abundant and best characterised SINE is the primate specific Alu element

(approximately 300bp) (Schmid and Jelinek, 1982). Truncated copies of these elements are often identified. Both of these elements are characteristic of retrotransposable elements which would explain their widespread distribution throughout the human genome (Britten *et al.*, 1988).

Tandemly repeated DNA. There is a large range of tandemly repeated DNA in the human genome both in repeat unit size and total array length. Satellite DNA has the largest array lengths constituting a large proportion (approximately 8%) of the human genome (Kit, 1961). At least five different classes of satellite DNA have been characterised (see Tyler-Smith and Brown, 1987). *In situ* hybridisation reveals that satellite DNA is located in constitutive heterochromatic regions (Jones and Corneo, 1971). The best characterised satellite DNAs are the α-satellites present at the centromeres of all human chromosomes (Manuelidis, 1978) and consist of a 171bp repeat unit arranged in long tandem arrays of several megabases in length (reviewed by Willard, 1990). These arrays constitute approximately 5% of the human genome and display variation in length and internal composition (Mahtani and Willard, 1990; Greig *et al.*, 1993). Other satellites include the β-satellites composed of large arrays of a 68bp repeat unit. These satellites constitute a large proportion of the short arms of the acrocentric chromosomes (Greig and Willard, 1992).

Midisatellite loci have repeat unit lengths smaller than that of satellite DNA and have array lengths of upto 500kb. Only few midisatellite loci have been characterised, for example, the D1Z2 locus which contains a tandem array of a 40bp repeat unit between 250-500kb long (Nakamura *et al.*, 1987). These loci are polymorphic in length and internal composition. For example, the midisatellite at DXYS20 consists of large arrays of upto tens of kilobases of a 61bp repeat unit, the arrays are highly variable in length and the repeat unit sequence is heterogenous with upto a 5% divergence between repeat units (Page *et al.*, 1987).

Minisatellite loci have repeat unit sizes in the range of 9bp to 90bp (Wong *et al.*, 1987; Armour *et al.*, 1990) tandemly repeated in arrays of between 0.5kb and 30kb long and there are an estimated 1500 such loci in the human genome (Jeffreys, 1987). Minisatellite loci are the most variable in the human genome with the heterozygosities at some minisatellites in excess of 90% (Wong *et al.*, 1987). *In situ* hybridisation revealed that approximately 80% of variable minisatellite loci are located in the proterminal regions of human chromosomes (Royle *et al.*, 1988) and the analysis of genetic linkage maps also revealed a clustering of these loci towards the end of the chromosomes (NIH/CEPH, 1992). The hypervariability at some minisatellite loci must be due to a high underlying mutation rate and it is at these loci that the most detailed knowledge of repeat unit turnover is available.

In addition to the variation in length, some minisatellites contain internal sequence variation between repeat units (Jeffreys *et al.*, 1991). It is possible to assay the distribution of the different repeat unit types within the minisatellite repeat array using a PCR-based technique called minisatellite variant repeat mapping by PCR (MVR-PCR) (Jeffreys *et al.*, 1991). This produces a digital code corresponding to the position of the different repeat types within the array. MVR-PCR revealed an even greater level of variation than that detected by allele length analysis and the majority of the variation detected was localised to one end of the repeat array (Jeffreys *et al.*, 1991; Armour *et al.*, 1993; Neil and Jeffreys, 1993). Analysis of mutations isolated from pedigrees or directly from sperm revealed that these mutation events were complex and often involved gene-conversion like events (Jeffreys *et al.*, 1994). This implied that elements outside of the repeat array mediate the mutation processes and later variants were detected in the DNA flanking the array which where associated with a reduction in minisatellite mutation (Monckton *et al.*, 1994).

The shortest tandemly repeated DNA sequences are simple reiterated mononucleotides often poly(dA) associated with the tails of retrotransposons such as Alu elements (Economou *et al.*, 1990). Microsatellite loci consist of small repeat units between 2bp and 8bp in arrays generally no more than 1kb. Microsatellite loci are often polymorphic and they are widespread throughout the human genome. This coupled with the fact that they can easily be typed using the polymerase chain reaction (PCR) makes them of great utility in the generation of genetic maps. The variability of microsatellite loci depends upon the repeat unit length and the number of identical repeat units (Weber, 1990; Armour *et al.*, 1994). Dinucleotide repeats are probably the most abundant microsatellites and it is estimated that there are between 50,000 and 100,000 in the human genome (Miesfeld *et al.*, 1981). Trinucleotide repeats are of particular interest in view of the involvement of trinucleotide repeat array expansion in a number of human genetic diseases (Yu *et al.*, 1991; Brook *et al.*, 1992; Chung *et al.*, 1993).

This Work

Telomeres are biologically important structures involved in the maintenance of chromosomal stability. In recent years they have become the focus for attention with the discovery of a possible role in cellular ageing and tumourgenesis. To date, the majority of analyses of the dynamics of telomeres have been carried out by observing all telomeres simultaneously using Southern blot technology. Single telomeres have been analysed at naturally occurring loci (Cooke and Smith, 1986) or at genetically engineered loci (Murnane *et al.*, 1994). The aims of

the project described in this thesis were to develop a technique for the detailed analysis of a single telomere both at the human population level or in a limited number of cells for *in vitro* and *in vivo* analysis on senescent cell lines and tumours. This was achieved with a technique similar to MVR-PCR that has been so successful at revealing the true extent of allelic variation and the processes underlying the high mutation rate at minisatellite loci. The technique described in this thesis allows the detailed analysis of variation in the distribution of different telomere repeats within a single telomere repeat array to be carried out. The information obtained allows inferences to be made about the processes influencing the turnover of these sequences.

Chapter 2

MATERIALS AND METHODS

The specific reaction conditions of the majority of the techniques used in this thesis are described in the relevant sections of the thesis.

Materials

Chemical reagents, molecular biological reagents and equipment. Chemical and molecular biological reagents were obtained from recognised suppliers. Chemicals were obtained from Fisons (Loughborough, UK), Serva (Heidelberg, Germany), Sigma Biochemical company (Poole, UK) and FMC Bioproducts (Rockland, USA). Molecular biological reagents were obtained from Gibco-BRL (Paisley, UK), New England Biolabs (Hitichin, UK), Advanced Biotechnologies (Leatherhead, UK), Pharmacia (Milton Keynes, UK), Boehringer Mannheim, (Lewes, UK) and Bio-RAD (Watford, UK). Radiochemicals were obtained from Amersham International plc (Little Chalfont, UK). Specialised equipment was obtained from Bio-RAD, Perkin-Elmer/Applied Biosytems (Beaconsfield, UK) and Hybaid (Teddington, UK).

Oligonucleotides. Oligonucleotides were synthesised by D. Langton (Protein and Nucleic Acids Laboratory, University of Leicester) and J. Keyte (Department of Biochemistry, University of Nottingham.). Oligonucleotides were ethanol precipitated and resuspended in distilled water, and the concentration of the oligonucleotide calculated from the absorbance at 260nm using a Cecil Instruments CE 202 Ultraviolet Spectrophotometer. Fluorescently labelled oligonucleotides were synthesised by Oswell DNA (University of Southampton) and purified by high pressure liquid chromatography (HPLC). The sequences and annealing temperatures of all the oligonucleotides used are shown in table 2.1. Random hexadeoxyribonucleotides for random primed labelling reactions were supplied by Pharmacia.

Human DNAs. DNAs obtained from the lymphoblastoid cell lines derived from kindreds which comprise the Centre d'Etude du Polymorphism Humain (CEPH, Paris, France) panel of Caucasian DNAs were supplied by Professor H. Cann and Dr. J. Dausset. Other Caucasian DNAs were obtained from venous blood samples taken from members of the Department of Genetics (University of Leicester, UK) by Dr. J. A. L. Armour. Blood and DNA samples from unrelated Japanese individuals were kindly donated by Professor Y. Katsumata (Nagoya University,

Table 2.1. PCR primer sequences and annealing temperatures.

PCR primer	Sequence (5′-3′)	Annealing temperature (°C)
TSK8A	GAGTGAAAGAACGAAGCTTCC	55
TSK8B	CCCTCTGAAAGTGGACCTAT	65
TSK8C	GCGGTACCAGGGACCGGGACAAATAGAC	65
TSK8E	GCGGTACCTAGGGGTTGTCTCAGGGTCC	65
TSK8F	GCGGATCCTACTCTCTATGGGCCACAC	60
TSK8G	CGGAATTCCAGACACACTAGGACCCTGA	65
TSK8H	GTACAAAAAAGCAATCACCAGATCAGC	60
TSK8J	GAATTCCTGGGGACTGCGGATG	65
TSK8K	CATCCCTGAAGAAGCATCTTGGCC	65
primA	GCGGTACCCAAAGACAGAAGGCCCAG	66
primB	GCGGTACCGAACTGGGTTATCGACCAGGT	66
Cht7B	CCGCTATCTGTATAAACATGGAAA	66
Telomere repeat primers		
TAG-TelC	TCATGCGTCCATGGTCCGGACTAACCCTAACCCTAACCCTAA	67
TAG-TelD	TCATGCGTCCATGGTCCGGACCTCACCCTCACCCTCACCCTCA	67
TelC	GGCCATCGATGAATTCTAACCCTAACCCTAACCCTAA	67
TelD	GGCCATCGATGAATTCCTCACCCTCACCCTCACCCTCA	67
TelG	CCCTCACCCTCACCCTCACCCTC	67
TelH	CCCTAACCCTAACCCTAACCCTA	67
TelF	CCCIIACCCIIACCC(CT)(AC)ACCCTC	67
TelL	ACCIIACCCIIACCCIIACCCTC	67
TelK	CCCTGACCCTGACCCTGACCTG	67
TelO	CCCTIACCCTIACCCT(ACGT)ACCCTA	67
TelP	CCCTIACCCTIACCCT(ACGT)ACCCTG	67
TelR	CCCTIACCCTIACCCT(ACGT)ACCCTC	67
TelT	CCCTIACCCTIACCCT(ACGT)ACCCTT	67
TelW	CCCTTACCCTTACCCT(ACGT)ACCCTA	67
TelX	CCCTTACCCTTACCCT(ACGT)ACCCTC	67
TelY	CCCTTACCCTTACCCT(ACGT)ACCCTG	67
TAG-TelJ	TCATGCGTCCATGGTCCGGAACCCCAACCCCAACCCCAACCCC	67
TAG-TelW	TCATGCGTCCATGGTCCGGACCCTTACCCTTACCCTNACCCTA	67
TAG-TelLX	TCATGCGTCCATGGTCCGGACCCTTACCCTTACCCTNACCCTC	67
TAG-TelY	TCATGCGTCCATGGTCCGGACCCTTACCCTTACCCTNACCCTG	67
Allele Specific Primers		
TS-842C	AGACGGGGACTCCCGAGC	67.5
TS-30A	CTGCTTTTATTCTCTAATCTGCTCCCA	68
TS-30T	CTGCTTTTATTCTCTAATCTGCTCCCT	67
TS-30TS	CTTTTATTCTCTAATCTGCTCCCT	62.5
TS-13AR	ACCCTCTGAAAGTGGACCA	66

Where two primers are used that have different annealing temperatures the lower of the two annealing temperatures is used, except for the allele specific primers where the temperature indicated for those primers is used. Bold letters indicate non-complementary 5′ tails and the underlined sequence represents restriction enzyme sites.

Japan). DNA samples from West-Africans were collected from Ibadan, Nigeria, the majority of the population in this area are members of the Yoruba tribe. The Afro-Caribbean DNA samples were provided by Cellmark Diagnostics, Abingdon, UK. These Afro-Caribbean sample were obtained from individuals involved in paternity cases in the London area; racial identification was made from photographs of the donors. DNAs from unrelated Zimbabwean individuals were obtained from sperm samples kindly supplied by A. D. Nkomo and S. B. Kanoyangwa (Forensic Science Laboratory, Causeway, Zimbabwe). These samples were obtained from individuals in the Harare area and are most likely to be from members of the Shona tribe. DNA samples from members of the Amazonian Karitiana tribe were kindly donated by Dr. K. K. Kidd.

Methods

DNA amplification. DNA was amplified using the polymerase chain reaction (PCR)(Saiki *et al.*, 1986) and unless otherwise stated, was carried out using approximately 100ng of genomic DNA, 1μM of each primer, 0.1 U/μl of Taq polymerase (Advanced Biotechnologies) and a PCR buffer. The standard PCR buffer was kept as an 11×stock and used at 1×concentration which consisted of 45mM Tris-HCl (pH8.8), 11mM $(NH_4)_2SO_4$, 4.5mM $MgCl_2$, 6.7mM β-mercaptoethanol, 4.4μM EDTA, 1mM dATP, 1mM dCTP, 1mM dGTP, 1mM dTTP and 113μg/ml BSA. The PCR reactions were cycled using a Geneamp 9600™ thermal cycler (Perkin Elmer) or with a 4800™ thermal cycler (Perkin Elmer) overlaid with a small drop of paraffin oil. Where possible PCR amplifications were setup in a Category II laminar flow hood for user and sample protection and care was always taken to keep amplified products separate from all PCR reagents, consumables and tools used during manipulations. Zero controls were always included.

Agarose gel electrophoresis. Agarose gel electrophoresis was carried out using horizontal submarine agarose gels containing ethidium bromide (0.5μg/ml) in 0.5×TBE (45mM Tris-borate, 1mM EDTA), using electrophoresis gel tanks manufactured in house and Bio-RAD electrophoresis power supplies. The DNA was visualised in the gel on a UV transilluminator (Chomato-vue C-63, UV products Inc.). The type of agarose and concentration depended upon the size of the DNA fragments to be resolved.

Preparative gel electrophoresis

From agarose gels. DNA fragments were resolved by electrophoresis through an agarose gel and the fragment of interest was excised from the gel in an agarose block. The agarose block was placed in a well cut in a fresh agarose gel. A piece of dialysis membrane was cut to a size slightly bigger than the gel slice and placed around the gel slice, such that upon electrophoresis the DNA in the slice would be eluted onto the membrane. The elution of the DNA was monitored by observing the fluorescence of the ethidium bromide stained DNA with a hand held UV illuminator. Once the DNA was electroeluted onto the membrane (usually no more than 10 minutes), the membrane was quickly removed from the gel (with the current still on) and placed in a 1.5ml microfuge tube. The corner of the membrane was trapped in the lid of the tube and the solution containing the DNA separated by centrifugation (12000 rpm for 5 minutes) the membrane was washed with 20µl of distilled water followed by centrifugation. The solution containing the DNA was transferred to a fresh microfuge tube and the DNA recovered by ethanol precipitation.

From polyacrylamide gels. DNA fragments were also purified from dried polyacrylamide gels containing radioactively labelled PCR fragments. The band of interested was located on an autoradiograph and cut from the gel using a clean scalpel blade, the autoradiograph was oriented upon the gel by the use of radioactive marking. The dried polyacrylamide was separated from the backing paper and placed in 10µl of TE (10mM Tris, 1mM EDTA) and left at room temperature for several hours to allow the DNA elute from the polyacrylamide gel. 1µl of the DNA in solution was then reamplified by PCR with the same primers used to generate the PCR product detected in the polyacrylamide gel. The reamplified product was then purified for further analysis by preparative agarose gel electrophoresis.

Single Stranded Conformational Polymorphism analysis (SSCP). SSCP (Orita *et al.*, 1989) was carried out on PCR amplified DNA. PCR products were generated with the primers TSK8C and TSK8B, purified by preparative agarose gel electrophoresis and radioactively labelled radioactive PCR products generated by reamplification with the same primers using the standard PCR buffer modified such that the final concentration of each dNTP was 0.1mM. 20µl PCR reactions were carried out with 200ng of genomic DNA, the modified PCR buffer, 0.1U/µl Taq polymerase and 9.25kBq/µl [α-^{32}P]dCTP (Amersham). The reactions were cycled 29 times using a Geneamp 9600^{TM} thermal cycler (Perkin Elmer), each cycle consisted of denaturation at 96°C for 50 seconds, primer annealing at 62°C for 60 seconds and extension at 70°C for 120 seconds. The PCR products were precipitated

in one volume of isopropanol with 2.5M ammonium acetate and left on ice for 15 minutes, followed by centrifugation at 12,000 rpm for 10 minutes and the pellet washed in 80% ethanol. The pellet was dried and resuspended in 10µl of distilled water. Sequence variation can generally not be detected using SSCP in DNA fragments of over 300bp in length, this 900bp product was digested prior to SSCP. 4µl of the PCR products was digested in a 10µl volume with *MspI* and *HindIII* and another 4µl digested with *MspI* alone. Duplicates of 3µl of the digestion were taken and to each was added 3µl of a solution containing 95% formamide, 20mM EDTA, 0.05% xylene cyanol, 0.05% bromophenol blue. One of the duplicates was heat denatured at 90°C for 3 minutes immediately prior to loading upon a 4°C SSCP gel. The denatured and native DNA fragments from any one individual were loaded in adjacent lanes on the SSCP gel. The SSCP gel consisted of 6% (19:1) polyacrylamide/bisacrylamide gel (National Diagnostics, Atlanta, USA), 10% Glycerol and 1×TBE. The gel was cooled to 4°C and run in a 4°C cold room to prevent denaturation of any secondary structure the single stranded products adopted during the migration through the gel. Electrophoresis was carried out at approximately 30W for 4-5 hours. The gel was transferred to 3mm filter paper (Whatman) and dried in a Bio-RAD 583 gel drier. The products were detected by autoradiography.

Telomere variant repeat mapping by PCR (TVR-PCR).

Manual TVR-PCR. The development of TVR-PCR technique was progressive and the various different stages of development are described in the text of this thesis (see chapter 4). The optimal conditions for TVR-PCR using the primers TAG-TelW, TAG-TelX and TAG-TelY in separate reactions with either one of the allele-specific flanking primers TS-30A or TS-30T are described here. The flanking primer (TS-30A or TS-30T) was 5′ end labelled in a reaction containing 0.5µM of the primer, 50mM Tris-HCl (pH 7.6), 10mM $MgCl_2$, 0.2U/µl of T4 polynucleotide kinase and 37kBq/µl [γ-^{32}P]ATP (Amersham). The 10µl PCRs contained 100ng of genomic DNA, 0.4µM of one of the telomere repeat primer (TAG-TelW, TAG-TelX and TAG-TelY) and 0.1µM of a radioactively 5′ end labelled allele-specific flanking primer, PCR buffer (1×) and 0.1 U/µl of Taq polymerase. The reactions were cycled 19 times using a Geneamp 9600^{TM} thermal cycler (Perkin Elmer), each cycle consisted of denaturation at 96°C for 20 seconds, primer annealing at 67°C for 40 seconds and extension at 70°C for 120 seconds. To each PCR, 5µl of a solution containing 95% formamide, 20mM EDTA, 0.05% xylene cyanol, 0.05% bromophenol blue was added. 6.5µl of the PCR reaction was electrophoresed through 6% (19:1) denaturing polyacrylamide/bisacrylamide gel (National Diagnostics, Atlanta, USA) containing 7.67M urea and a buffer gradient of 0.5-2.5×TBE until the xylene cyanol was approximately 5cm from the base of the gel. The gel was

transferred to 3mm filter paper (Whatman) and dried using a Bio-RAD 583 gel drier and the products detected by autoradiography with Fuji RX100 X-ray film. The gel apparatus used was a Sequi-Gen^{TM} 38cm×50cm IPC (Bio-RAD) with a sharks tooth comb with 6.3mm wide wells. Higher molecular weight fragments in the TVR-PCR patterns where further resolved by electrophoresis through a 5% (19:1) polyacrylamide/bisacrylamide gel containing 7.67M urea and 1×TBE for at least 7 hours.

Semi-automated TVR-PCR. The resolution and observation of TVR-PCR products was also carried out using an Applied Biosytems (ABI) 373. TVR-PCR reactions were carried using the TS-30A primer fluorescently labelled with one of the ABI dyes HEX, 6-FAM or TET (HPLC purified). The TS-30A primer labelled with one of the fluorescent dyes was used with one of the telomere repeat primers. The reaction and cycling conditions were the same as described for manual TVR-PCR except that 25 PCR cycles were required. The separate reactions were pooled; 2µl were combined with an equal volume of 100% formamide containing Dextran Blue (Dextran Blue is used as a colour indicator to aid in the loading of the sample, this dye does not enter the polyacrylamide gel); the samples were then denatured at 96°C for 2 minutes and placed upon ice. 2.5µl of the denatured samples where loaded onto a 5% (19:1) denaturing polyacrylamide/bisacrylamide gel containing 6M urea 1×TBE. The gel was prepared according to the manufactures instructions, taking special care to avoid contact of the reagents and equipment with any substance capable of fluorescence. The gel was subjected to electrophoresis in an ABI 373 automatic DNA sequencer running the gene-scan software 672 version 1.2.2. (Applied Biosytems).

Southern blot analysis. The DNA in agarose gels was depurinated (2×7mins in 0.25M HCl), alkali-denatured (2×15mins in 0.5M NaOH, 0.1 NaCl), then neutralised (30mins in 0.5M Tris-HCl (pH7.4), 3M NaCl). The DNA was transferred from the agarose gel to a Hybond-N FP (Amersham) nylon membrane by Southern blotting for at least two hours using 20×SSC as a transfer buffer. The nylon membrane was dried and the DNA covalently linked to it by exposure to 7×10^4 J/cm^2 of UV radiation in a ultraviolet crosslinker (Amersham). DNA sequences were detected by hybridisation to a radioactively labelled DNA probe. 10ng of the probe was labelled by the random priming method using randomly generated hexamers and DNA polymerase (Klenow fragment) (Fienberg and Vogelstein, 1984) to incorporate [α-^{32}P]dCTP (described in Wong et al., 1987). After stopping the 25µl labelling reaction with 70µl of 'stop' solution (20mM NaCl, 20mM Tris-HCl (pH 7.5), 2mM EDTA, 0.25% SDS), the probe was recovered by ethanol precipitation using 100µg of high molecular weight herring sperm DNA as a carrier molecule and

washed in 80% ethanol. The probe was then resuspended in 0.5ml of distilled water and denatured by boiling immediately prior to use. The Southern blots were prehybridised at 65°C for at least 20 minutes in the 10ml of the hybridisation solution between nylon meshes separating the membranes to allow even contact with the solution. Two different hybridisation solutions were used; a modified Church and Gilbert (1984) phosphate/SDS solution (0.5M sodium phosphate (pH 7.2), 7% SDS, 1mM EDTA) (Wong *et al.*, 1987); the other hybridisation solution included 1.5×SSPE (10×SSPE; 1.48M NaCl, 0.16M disodium hydrogen phosphate, 5mM EDTA, HCl to pH 7.7), 1% SDS, 0.5% Milk powder, 6.1% PEG 6000 and is commonly known as the 'blotto' mix. The prehybridisation solution was replaced with a fresh 10ml of the same solution to which was added 0.5ml of the boiled probe. After the Southern blot was left to hybridise at 65°C in a bottle in a hybridisation oven (Hybaid) over night. The hybridisation membranes were then washed at least three times in 1×SSC/0.1%SDS to 0.1×SSC/0.1%SDS depending upon the wash-stringency required. The hybridisation patterns were visualised by autoradiography with Fuji RX100 X-ray film. The exposure time and whether an intensifying screen was used for a -70°C exposure depended upon the strength of the hybridisation signal.

DNA sequencing. Good quality DNA sequence was generated from PCR amplified double stranded DNA purified by preparative gel electrophoresis. The dideoxy chain termination procedure used was that described by P. Winship (1989). This technique utilises di-methyl sulphoxide to alleviate problems occurring with the reannealing of the template strands. The products of the sequencing reaction were resolved by electrophoresis through a 6% (19:1) denaturing polyacrylamide/bisacrylamide 0.5-2.5×TBE buffer gradient gel containing 7.67M urea and visualised by autoradiography.

Restriction fragment length polymorphism (RFLP) assays.
Assays for the -1888 *NlaIV*, -842 *DdeI*, -652 *AvaII* and -415 *MboII* polymorphisms are described in the text.
The -176 (*DdeI*) polymorphism was assayed by the PCR amplification with the primer TSK8E(1μM) and a mixture of the TS-13AR(0.5μM) and TSK8B(0.5μM) primers to ensure amplification was obtained from both alleles. The 10μl PCRs contained the components described above and were cycled 32 times in a Geneamp 9600™ thermal cycler (Perkin Elmer), each cycle consisted of denaturation at 96°C for 20 seconds, primer annealing at 62°C for 40 seconds and extension at 70°C for 50 seconds. 5μl of the PCR product was digested with *DdeI* in a 16μl reaction and the products resolved by electrophoresis a 4.5% metaphor agarose gel (FMC). The 418bp PCR fragment was digested into many fragments, the presence of a DdeI site at the -176 position cleaves the 155bp fragment into a

114bp fragment and a 41bp. Haplotype analysis to include the -176 polymorphism was carried out by allele specific amplification from the -842 polymorphism in heterozygous DNAs, using the primer TS-842C(1μM) with a mixture of TSK8B(0.5μM) and TS-13AR(0.5μM) followed by digestion with *DdeI* and resolution of the products by electrophoresis through a 3.5% Metaphor agarose gel (FMC) in 0.5×TBE. Again the presence of a DdeI site at the -176 polymorphism results in the cleavage of the 155bp fragment into two fragments of 114bp and 41bp.

The -30 polymorphism was assayed using an amplification refractory mutation system (ARMS) system (Newton *et al.*, 1989) whereby sequence variants are detected using a PCR primer designed to specifically amplify single alleles. Two separate PCR amplifications were carried out, each containing an allele-specific primer (TS-30A(1μM) or TS-30TS(1μM)) in conjunction with a mixture of the primers TSK8B(0.5μM) and TS-13AR(0.5μM). To achieve allele specificity the reactions containing TS-30A were cycled 32 times, each cycle consisted of denaturation at 96°C for 15 seconds, primer annealing at 68°C for 40 seconds and extension at 70°C for 30 seconds; the reactions containing TS-30TS were cycled 33 times, each cycle consisted of denaturation at 96°C for 15 seconds, primer annealing at 62.5°C for 40 seconds and extension at 70°C for 30 seconds. The 60bp products were resolved on a 4.5% Metaphor agarose gel (FMC) in 0.5×TBE.

Computing. Computerised analysis of DNA sequences was carried out using the Genetics Computer Group (GCG) Sequence analysis software version 6.2 (Devereux *et al.*, 1984) running on a Scilicon Graphics Inc. 4D/480S system using IRIX release 5.3.

Photography. Photographs of ethidium bromide stained gels were generated by visualisation of the gel on a UV transilluminator (Chomato-vue C-63, UV products Inc.) followed by photography using a Polaroid MP-4 camera with Kodak negative film (T-max Professional 4052). The negatives and autoradiographs were developed using processing reagents recomended by Kodak. Prints of negatives and autoradiographs was carried out by the Central Photographic Unit (University of Leicester). Laser photocopies and colour prints outs were generated by the Central Reprographics Unit (University of Leicester).

Chapter 3

VARIATION IN THE DNA FLANKING THE Xp:Yp PSEUDOAUTOSOMAL TELOMERE

Summary

In this chapter I have described the detailed analysis of sequence variation in the DNA flanking the Xp:Yp pseudoautosomal telomere. This was initially carried out in order to obtain polymorphisms that could be utilised to access variation in the telomere array itself, but the analysis revealed some interesting and unexpected features of the telomere adjacent DNA. A high frequency of polymorphism was identified in the DNA immediately flanking the telomere. This consisted of a high frequency of single base substitutional polymorphisms and one 10bp insertion/deletion polymorphism. The frequency of polymorphism was one in every 68bp in Caucasians and one polymorphism in 46bp in West-Africans. Restriction fragment length polymorphism (RFLP) analysis revealed significant linkage disequilibrium. Over the seven polymorphic positions analysed spanning 1858bp, just seven of the possible 49 haplotypes that might be expected were detected in 78 unrelated Caucasian DNAs. Similar analysis in 30 unrelated Japanese DNAs revealed only four of the possible 49 haplotypes. A more limited analysis of 27 West-African DNAs also revealed strong but reduced linkage disequilibrium across five polymorphic positions spanning 812bp adjacent to the telomere, such that four of the possible 25 haplotypes were observed. Sequence comparison of 846bp immediately flanking the telomere in the different haplotypes found in Caucasian and West-African individuals, revealed a maximum divergence of 1.5% and 2.2% respectively. The extent of the sequence divergence between the haplotypes suggests that they are as ancient as the divergence between the chimpanzee (*Pan*) and human (*Homo*) lineages.

Introduction

The Xp:Yp pseudoautosomal pairing region. The Xp:Yp pseudoautosomal pairing region is a 2.6Mbp region of homology between the X and Y chromosomes. The proximal limit of this region is defined by an Alu element (Ellis *et al.*, 1989) and the distal limit by the telomere (Cooke *et al.*, 1985). There is a gradient of sex linkage across the Xp:Yp pairing region such that genetic markers near the telomere exhibit a 'pseudoautosomal' mode of inheritance (Rouyer *et al.*, 1986). Chromosome pairing in the pseudoautosomal region and the formation of an obligate recombination is thought to be essential for the correct segregation of these chromosomes during male meiosis (Cooke *et al.*, 1985; Rouyer *et al.*, 1986; Shapiro *et al.*, 1989). Recombination in the pseudoautosomal region is approximately 10-fold higher in males (50 cM) than in females (4-18 cM) (Rouyer *et al.*, 1986; Page *et al.*, 1987). This difference is thought to be due to the fact that, in females, there is opportunity for genetic exchange along the entire length of the X chromosome. This results in a relative reduction in the likelihood of exchange in the pseudoautosomal region (Soriano *et al.*, 1987). This could also be due to differences in pairing and/or recombination of the pseudoautosomal region between male and female meiosis (Hunt and LeMaire, 1992). The telomeres of the Xq and Yq chromosome arms can associate during meiosis (Speed and Chandley, 1990). Molecular cloning of the terminal region of the Xq and Yq chromosome arms revealed that these chromosomes share 400kb of sequence and the analysis of genetic markers in this region revealed non-obligate recombination in male meiosis, this defines a second pseudoautosomal region (Freije *et al.*, 1992).

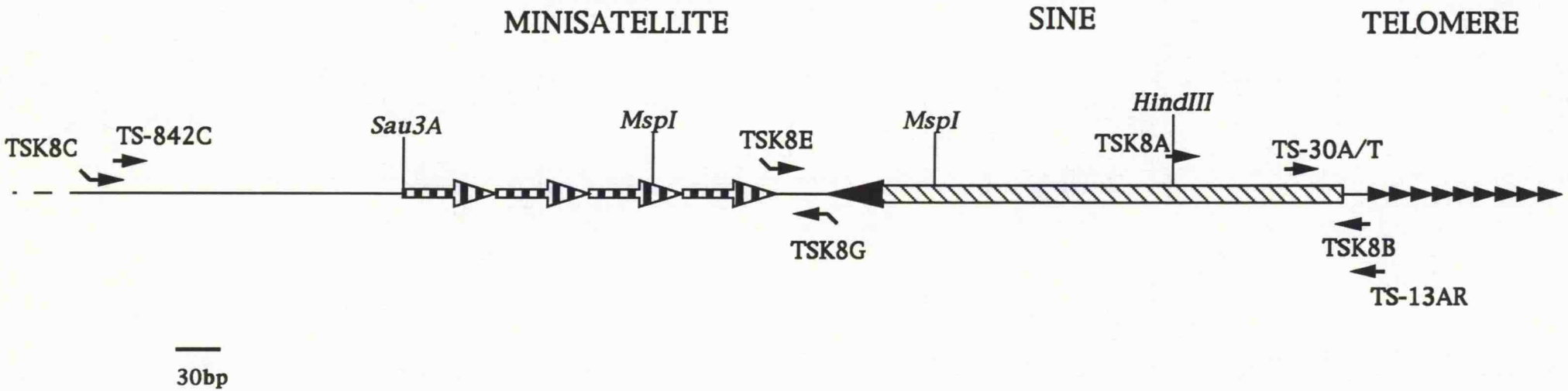
The DXYS14 locus. This locus, situated within 13.4kb of the start of the telomere repeat array, was isolated during a screen for CpG-rich sequences on the Y chromosome and it is detected by the clone 29C1 (Cooke *et al.*, 1985; Brown, 1989). The end of the genetic map of the Xp:Yp pseudoautosomal region is defined by DXYS14, which consists of a hypervariable minisatellite locus with a heterozygosity of 97% and an overall mutation rate of the order of 1.3% per gamete (Inglehearn and Cooke, 1990). This locus has sites for *SstII*, *HpaII* and *HhaI* and a G+C content of 68% which can be indicative of a low methylation CpG island associated with the 5′ end of genes (Lindsay and Bird, 1987; Inglehearn and Cooke, 1990). However, this locus is methylated in DNA extracted from blood and does not hybridise to RNA transcripts from a HeLa cell extract, suggesting that this locus may not be a CpG island.

The Xp:Yp pseudoautosomal telomere-junction region. The DNA immediately adjacent to the Xp:Yp telomere was originally isolated by a telomere complementation strategy utilising 'half-YACs' (Brown, 1989). It was demonstrated to be a unique chromosome end by *in situ* hybridisation to metaphase chromosomes. Approximately 2kb of DNA immediately flanking the telomere array has been sequenced (Brown *et al.*, 1990). The locus consists of a monomorphic single locus minisatellite containing four 64bp repeat units and a shortened version of a short interspersed nuclear element (SINE) 14 base pairs (bp) from the start of the telomere (Royle *et al.*, 1992). There are an estimated 20,000 copies of this family of SINEs within the human genome (Fig. 3.1.)(La Mantia *et al.*, 1989). Hybridisation with the minisatellite to genomic DNA digested with *Sau3A* revealed a smear of fragments, characteristic of a terminal restriction fragment (TRF). This locus was also isolated using a telomere-anchored polymerase chain-reaction approach (Royle *et al.*, 1992).

Characterising variation in the Xp:Yp telomere-junction region. The reasons for characterising variation at this locus are two fold. Firstly, an initial survey of variation at this locus using single stranded conformational polymorphism analysis (SSCP) revealed that there was a large amount of sequence variation (K. A. Duffy pers. comm.). This sequence polymorphism could be utilised to generate an informative terminal marker system. Comparison of genetic and physical maps demonstrates linkage map expansion in the proterminal regions of human chromosomes (NIH/CEPH *et al.*, 1992; Litt *et al.*, 1995), also, chiasmata tend to be distributed in these regions (Hulten, 1974). These data suggest that the proterminal regions of human chromosomes have elevated levels of recombination. The Xp:Yp pseudoautosomal region has a gradient of recombination such that markers towards the end of the chromosome recombine with a higher frequency than those situated in more proximal locations (Cooke *et al.*, 1985; Rouyer *et al.*, 1986; Page *et al.*, 1987). A terminal marker system could be used to estimate the number of different chromosomal ends in the population and provide information about recombination rates at the terminus of the Xp and Yp chromosome arms between the most distal locus on the genetic map DXYS14 and the telomere. Secondly, sequence polymorphism in the DNA flanking the telomere array could be utilised to analyse single telomere repeat arrays by allele specific amplification into the telomere in individuals heterozygous for telomere-adjacent polymorphisms (see Chapter 4).

This work. In this chapter I have described the detailed analysis of sequence polymorphism in the DNA immediately flanking the Xp:Yp pseudoautosomal telomere. The information obtained is considered from the biological standpoints of allelic diversity, mutation and evolution.

Fig. 3.1.



Xp/Yp PSEUDOAUTOSOMAL-TELOMERE JUNCTION

Results

1. Characterising variation

Single Stranded Conformational Polymorphism Analysis. SSCP is a powerful technique for detecting sequence variation. Briefly, it involves electrophoresis of labelled denatured DNA typically no longer than 300bp through a non-denaturing gel matrix. Differences in sequence are detected as the single strands of DNA exhibit differing mobilities in the gel matrix. This is due to the different conformations that the single stranded DNA molecules adopt (see chapter 2)(Orita *et al.*, 1989). SSCP has been widely used for the detection of single base substitution and insertion/deletion (ins/del) polymorphisms and mutations (Neil and Jeffreys, 1993; Cullen *et al.*, 1995).

I utilised this technique to investigate sequence polymorphism in the Xp:Yp pseudoautosomal telomere-junction region. Radioactively labelled Polymerase Chain Reaction (PCR) products were generated by amplification with primers TSK8C and TSK8B (Fig. 3.1.) in the presence of ^{32}P dCTP. PCR products from 35 unrelated Caucasian individuals (CEPH) were divided into two and digested with *MspI* alone or with *HindIII* and subjected to SSCP. This resulted in a screen for polymorphisms in the 480bp proximal to the start of the telomere repeat array. A remarkable amount of variation was detected by SSCP, so much that it was difficult to quantify. To determine the nature of this variation, direct sequence analysis was carried out.

Sequence Analysis. PCR products containing the SINE and the minisatellite were generated with primers TSK8B, TSK8E and TSK8C, TSK8G from 32 Caucasian (CEPH) and 21 West-African individuals. In Caucasians and West-Africans this would give a 97% [$100(1-0.9^{32})$] and a 89% [$100(1-0.9^{21})$] chance respectively of detecting a polymorphism with a frequency of 0.1 and a heterozygosity of 18% assuming Hardy-Weinberg equilibrium. The PCR products were purified by agarose gel electrophoresis followed by electroelution on to a dialysis membrane. Direct sequencing was carried out on the purified PCR products (see chapter 2) (Winship, 1989). The sequence analysis covered approximately 820bp (from -26 to -846). The numbers are distances in bp from the start of the telomere repeat array; the base -1 is the first base that is not in the telomere repeat array (Fig. 3.2b.). The bases -377 to -412 covered by the primers TSK8E and TSK8G were not screened by sequence analysis, but these primers have been used extensively at high annealing temperatures and no null amplifications have been observed, suggesting that no sequence polymorphism occurs in this 35bp region.

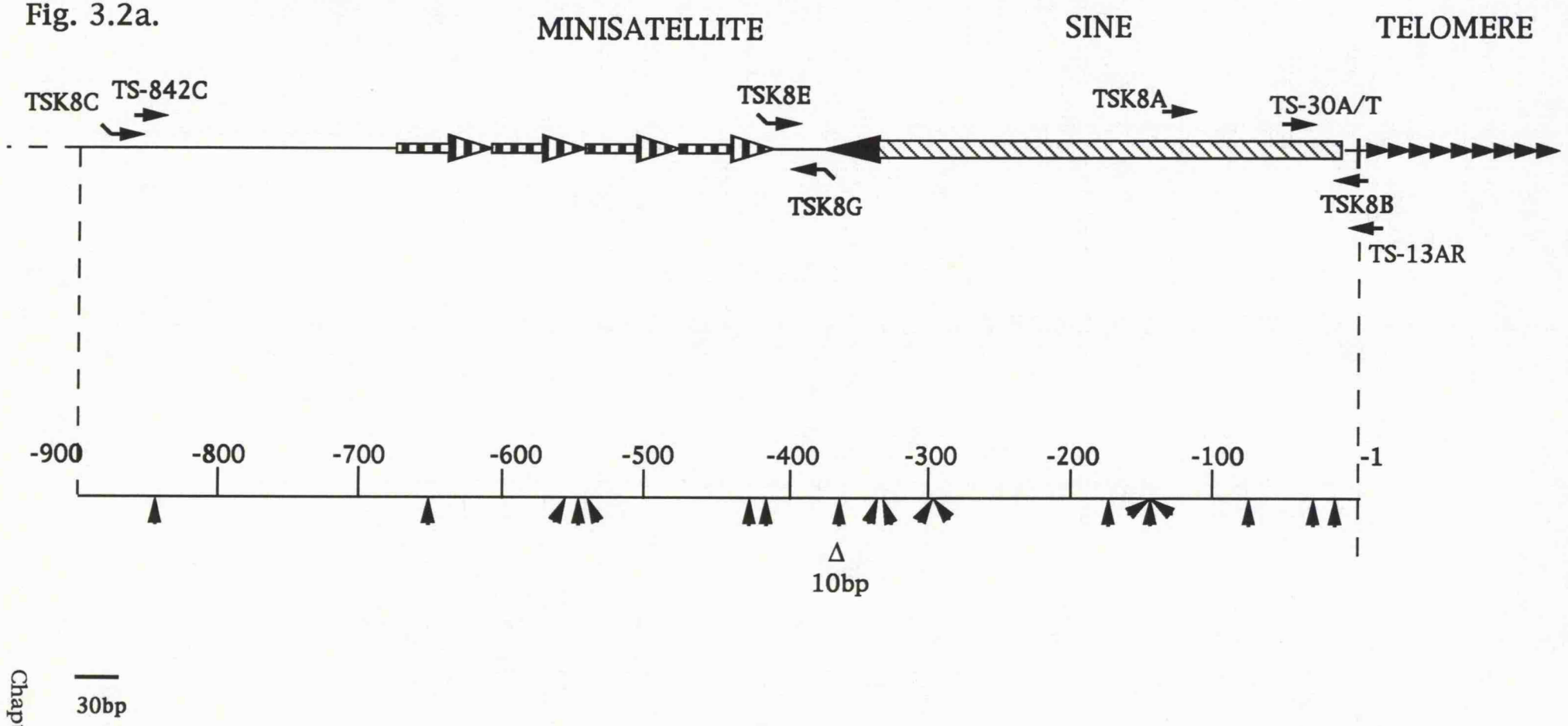
Fig. 3.2. Showing the polymorphism detected in the DNA flanking the Xp:Yp telomere

Sequence analysis of unrelated Caucasian and West-African DNAs revealed twelve and seventeen single base substitutional polymorphisms respectively in addition a 10 bp insertion/deletion polymorphism was detected in West-Africans only.

3.2a Shows the positions of the polymorphisms as ▲ below the line

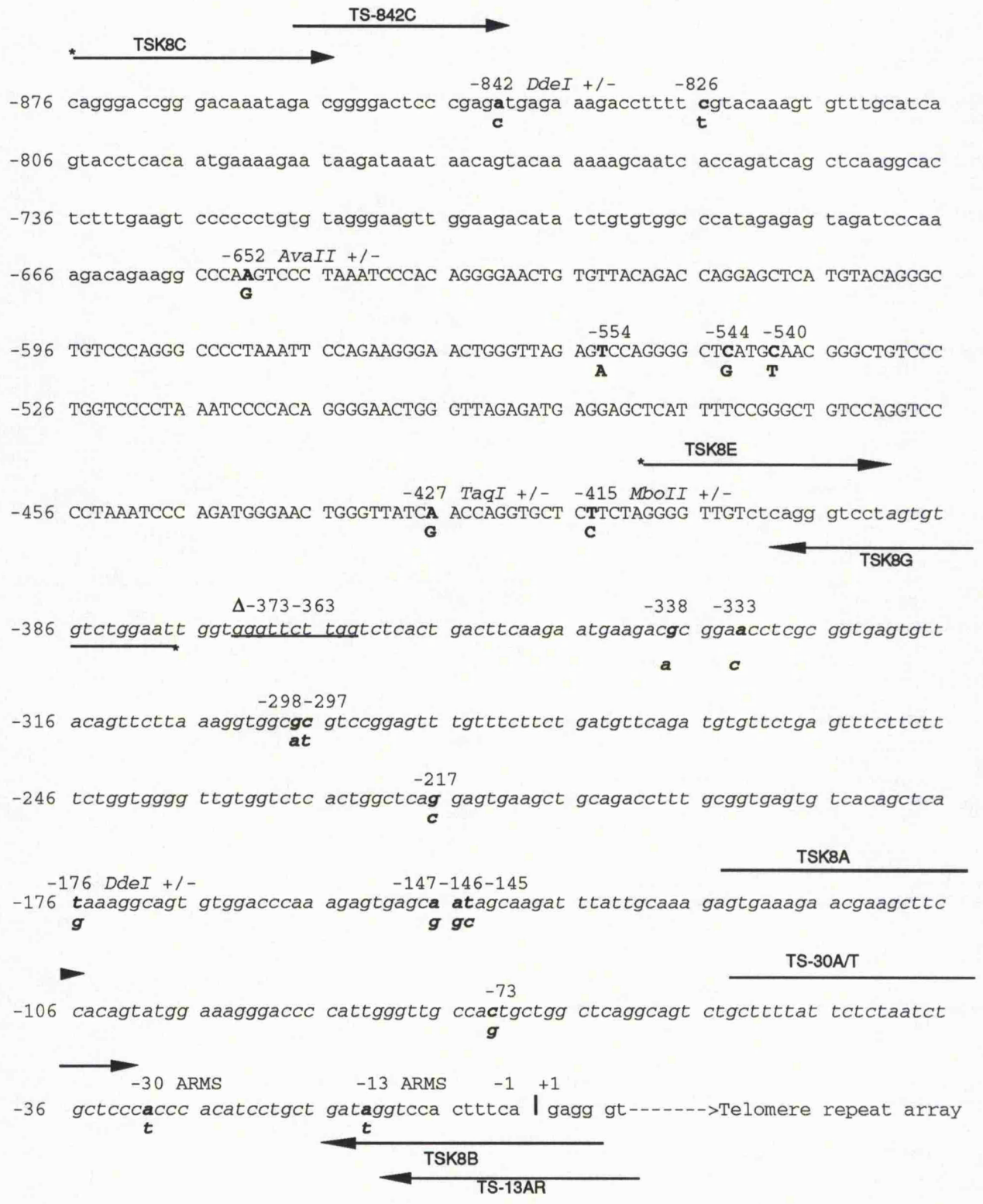
3.2b The sequence of the DNA flanking the Xp:Yp telomere. The polymorphisms detected are shown in bold. The PCR primers used during the analysis of this region are detailed above the sequence, an asterisk at the 5′ end of a primer denotes the presence of a non-complementary tail containing a restriction endonuclease site. The numbering system starts at beginning of the telomere repeat array as shown, positions proximal to this point are negative and positions distal are positive.

Fig. 3.2a.



Xp/Yp PSEUDOAUTOSOMAL-TELOMERE JUNCTION

Fig. 3.2b. Sequence of the 876 bp immediately flanking the Xp:Yp telomere with the polymorphisms detected and PCR primers.



Sequence analysis revealed a large number of single base substitutional polymorphisms. Twelve polymorphisms were identified in Caucasian and seventeen in West-African DNAs. In addition, a 10bp ins/del polymorphism was only found among African DNAs (Fig. 3.2a. and 3.2b.). This represents a frequency of one polymorphism every 68bp in Caucasians and one polymorphism every 46bp (including the 10bp ins/del polymorphism) in West-Africans. The polymorphisms identified are distributed throughout the 820bp sequenced and consist of ten transitions and seven transversions. Five of the polymorphisms have arisen at CpG doublets presumably by mutation of the C-residue by a methylation and deamination repair pathway (Fig 3.2b.).

2. Analysis of the -842, -652, -427, -414, -176, -30 and -13 polymorphisms in the Caucasian population.

Restriction Fragment Length Polymorphism Analysis. Some of the polymorphisms identified create or destroy a restriction enzyme site. These sites can therefore be assayed by PCR amplification followed by digestion with the appropriate diagnostic restriction enzyme and resolution by agarose gel electrophoresis (RFLP analysis). Four polymorphic sites were chosen for analysis; these were at -414 C-T *MboII*, -427 G-A *TaqI*, -652 A-G *AvaII* and -842 C-A *DdeI*. These all occur on the same PCR fragment (TSK8G-TSK8C) therefore only one amplification is required to assay all four sites. The PCR products were divided between four restriction enzyme digestions, one for each of the four sites. The products of the digestions are resolved on a 2.2% Metaphor agarose gel and observed by ethidium bromide staining (Fig. 3.3.). The survey was initially carried out on unrelated Caucasian (CEPH) DNAs (n=80). Each of the polymorphic positions analysed were in Hardy Weinberg equilibrium ($p>0.05$) and positions -414, -652 and -842 were present in the population with identical frequencies (Table 3.1.). The frequencies of the -427 polymorphism are shown in Table 3.2. Surprisingly, most individuals homozygous at one position were homozygous for the other three positions analysed; likewise, individuals heterozygous at one position were heterozygous at the other three (Table 3.3.). Therefore, it seemed that only a limited number of the 16 possible haplotypes were present in Caucasians.

Haplotype analysis. In order to determine the haplotypes present among the individuals heterozygous at the four polymorphic positions, a PCR primer was designed to specifically amplify from alleles with a C at the -842 position (TS-842C) (Fig. 3.2b.). All the Caucasian DNAs heterozygous at the -842 position were amplified with TS-842C and TSK8G, and the resulting PCR products were digested

Fig. 3.3. RFLP assays for polymorphisms flanking the Xp:Yp telomere in five unrelated individuals

All the assays are performed on PCR products of 498bp generated from the amplification of genomic DNA with primers TSK8G and TSK8C, followed by digestion with the appropriate diagnostic restriction enzyme.

A. -414 *MboII* RFLP. The presence of a C at this position creates a *MboII* site cutting the 314bp fragment into a 268bp and a 46bp fragments (the 46bp fragment is not shown on this gel), the 184bp fragment is constant.

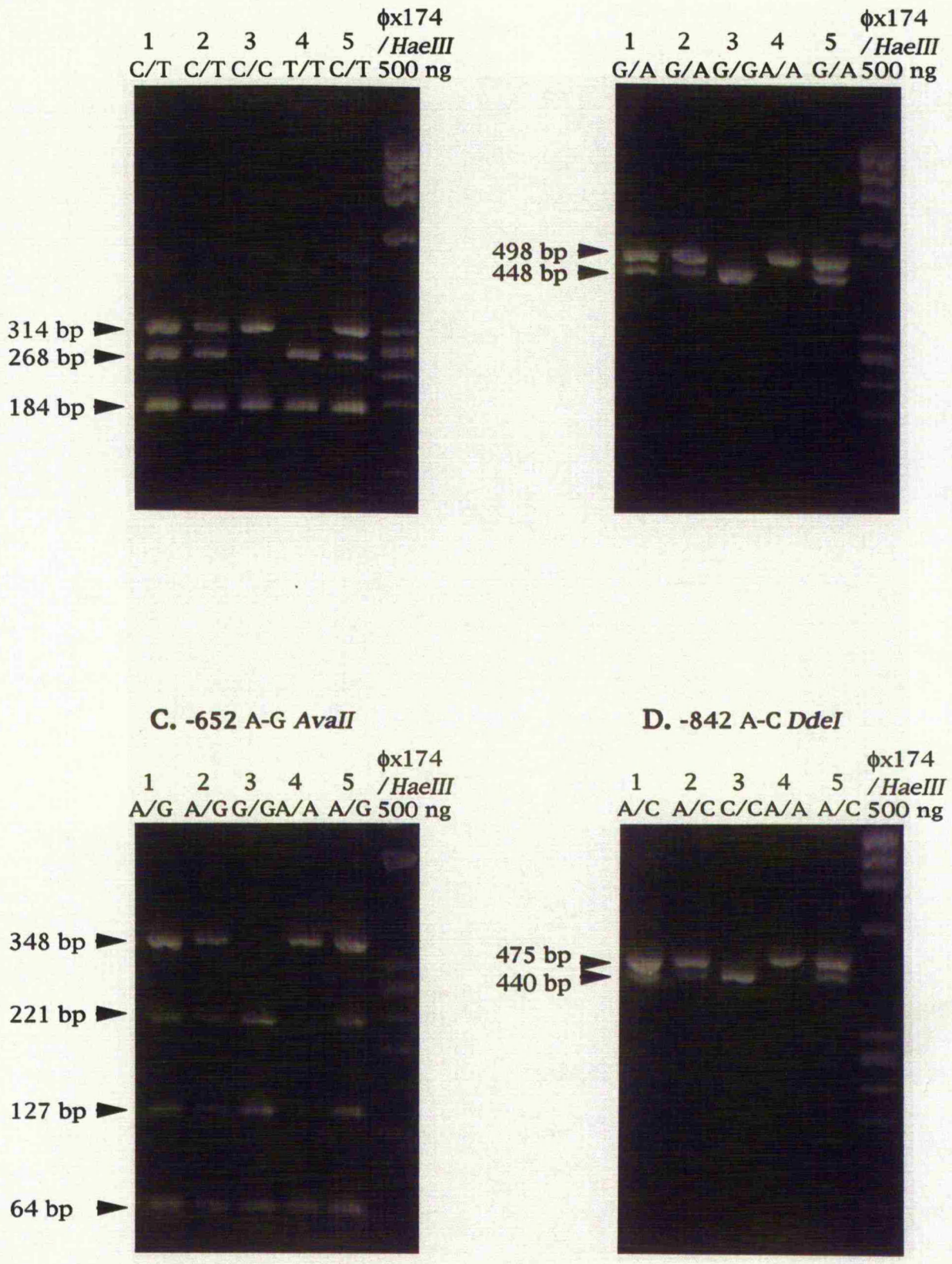
B. -427 *TaqI* RFLP. The presence of a G at this position creates a *TaqI* site cutting the 498bp PCR product to create two fragments of 448bp and 50bp (the 50bp fragment is not shown on this gel).

C. -652 *AvaII* RFLP. The presence of a G at this position creates an *AvaII* site cutting the 348bp fragment in to a 221bp and a 127bp fragment, the 64bp band is a doublet and is monomorphic, a 22bp monomorphic fragment is also released but is not shown on this gel.

D. -842 *DdeI* RFLP. The presence of a C at this position creates a *DdeI* site cutting the 475bp fragment into a 440bp and a 35bp fragment (the 35bp fragment is not shown on this gel) a 23bp monomorphic fragment is also released but is not shown on this gel.

The individuals shown in each assay are the same. This demonstrates the linkage disequilibrium observed, for example, individual 3 is homozygous at all four positions assayed.

The 20µl PCRs contained 1µM of each primer, the standard PCR buffer (see chapter 2) and 0.1 U/µl of Taq polymerase. The reactions were cycled 32 times using a Geneamp 9600^{TM} thermal cycler (Perkin Elmer), each cycle consisted of denaturation at 96°C for 20 seconds, primer annealing at 65°C for 40 seconds and extension at 70°C for 50 seconds. The PCRs were split equally into four restriction enzyme digests using the appropriate buffer and conditions recommended by the manufacture. The products of the reactions were resolved by electrophoresis through a 2.2% metaphor agarose gel (FMC) in 0.5×TBE and the fragments observed by staining the gel with ethidium bromide.



Chapter 3. Figure 3

Table 3.1. Allele frequencies at the -414, -652 and and -842 polymorphisms in 80 unrelated Caucasian individuals

Genotype at -414, -652, -842	Observed	Expected
C/C, G/G, C/C	22	21
C/T, G/A, A/C	37	40
T/T, A/A, A/A	21	20
Total	80	
Frequencies		
C, G, C	0.51	
T, A, A	0.49	
$\chi 2=$	0.45	P= 0.80

Table 3.2. Allele frequencies at the -427 polymorphism in 80 unrelated Caucasian individuals

Genotype at -427	Observed	Expected
G/G	26	25
G/A	37	39
A/A	17	16
Total	80	
Frequencies		
G	0.56	
A	0.44	
$\chi 2=$	0.32	P= 0.85

Table 3.3. Genotypes of 8 unrelated Caucasian individuals

CEPH	-414 *Mbo II*	-427 *Taq I*	-652 *Ava II*	-842 *Dde I*
201	C	G	G	C
202	C/T	G/A	G/A	A/C
1201	C/T	G/A	G/A	A/C
1202	C/T	G	G/A	A/C
1701	C	G	G	C
1702	C/T	G/A	G/A	A/C
2101	T	A	A	A
2102	T	A	A	A

with the appropriate diagnostic restriction enzyme for sites -652, -427 and -414. The products were observed after electrophoresis and ethidium bromide staining. All individuals heterozygous at all four positions shared two identical haplotypes which were the same as those observed in homozygous individuals (Fig. 3.4.). The haplotype analysis of 80 Caucasian individuals (160 alleles) revealed just three of the sixteen possible haplotypes across the four polymorphic positions. This demonstrated the existence of nearly complete linkage disequilibrium across this region. The three haplotypes identified were designated A, B and C. Haplotypes A and B differ at all four positions and haplotypes B and C differ at just one (-427) (Table 3.4a.). The haplotypes A, B and C are in Hardy Weinberg equilibrium in the Caucasian population ($p>0.05$) (Table 3.4b.). Linkage disequilibrium can be estimated using the disequilibrium coefficient which is defined as $D_{uv} = p_{uv} - p_u p_v$, where p_{uv} is the frequency of haplotype uv formed from alleles u and v at two loci with frequencies p_u and p_v respectively. This coefficient has values that extend from 0.25 for complete linkage disequilibrium to -0.25 for linkage equilibrium. An overall test of linkage disequilibrium can be calculated with the non-central chi-squared statistic $\sum_T^2 = \sum_{U=1}^{k}\sum_{v=1}^{l}\frac{nD_{uv}^2}{\tilde{p}_u\tilde{p}_v}$ which has $(k-1)(l-1)$ degrees of freedom when two loci have k and l alleles (chi-squared statistics were calculated using a computer program written Y. E. Dubrova in Microsoft Quickbasic^{TM} according to Weir, 1990). This analysis has been carried out on data for the -842, -652, -427 and -414 positions and demonstrates significant linkage disequilibrium in Caucasians across the -842, -652 and -414 positions ($\chi^2=80.00$, $p<0.01$), with reduced disequilibrium when the -427 position is included ($\chi^2=65.44$, $p<0.01$)(Table 3.4a.)

Identification of the -13 T-A polymorphism. To determine whether the linkage disequilibrium observed across the -842, -652, -427 and -414 polymorphic positions extends to other more distal polymorphisms, the -176 G-T *DdeI*, -30 T-A and -13 T-A polymorphic positions (Fig. 3.2.) were analysed. The -13 polymorphism was not detected during the initial screen. However, while developing an RFLP assay for the -176 position (Fig. 3.2b. and see below) it was clear that some alleles were not amplifying in the PCR reaction with the same efficiency. This could be due to additional polymorphism in either one of the primers used for the -176 assay (TSK8E or TSK8B) (Fig. 3.2b.). Sequence analysis was carried out in the region covered by the TSK8B primer and into the telomere. PCR amplifications from three DNAs with the telomere repeat primer TelH (which anneals to TTAGGG repeats) and TSK8E were generated. Products of approximately 500bp were purified and sequenced with primer TSK8A (Fig. 3.1.). An A-T transversion was detected 13 bases from the start of the telomere repeat array. This polymorphism is 2 bases from the 3′ end of the TSK8B primer, which would

Fig. 3.4. Haplotype analysis for polymorphisms flanking the Xp:Yp telomere

Haplotype analysis carried out on seven DNAs heterozygous for the -842, -652, -427 and -414 polymorphic positions. Amplification is carried out from genomic DNA from individuals heterozygous at the -842 polymorphic position using the allele specific primer TS-842C and TSK8G followed by digestion of the 480bp fragment with the appropriate diagnostic restriction enzyme.

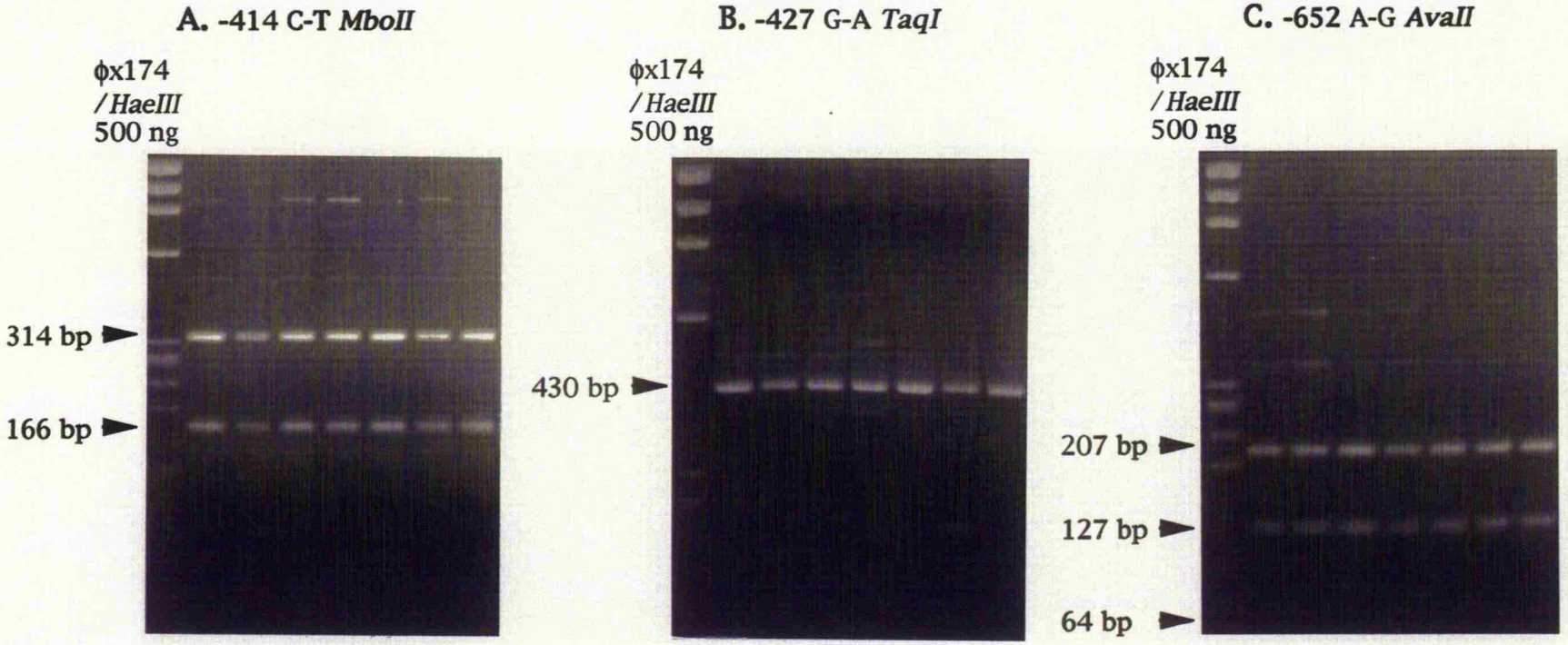
A. -414 *MboII* RFLP. Allele specific amplification and digestion with *MboII* results in the 314 bp fragment which correlates with the absence of the *MboII* site and therefore the presence of T at this position. The 166 bp fragment is monomorphic.

B. -427 *TaqI* RFLP. Allele specific amplification and digestion with *TaqI* results in a 430 bp fragment and a 50bp fragment which correlates with the presence of the *TaqI* site and therefore a G at this position (the 50bp fragment is not shown on this gel).

C. -652 *AvaII* RFLP. Allele specific amplification and digestion with *AvaII* reveal the 207 bp and the 127 bp fragments which correlates with the presence of the AvaII site and therefore a G at this position. The 64 bp band is a monomorphic doublet.

Allele specific amplification from the -842 position followed by digestion with the appropriate diagnostic restriction enzyme results in identical restriction patterns in all the individuals analysed. This demonstrates the existence of a limited number of haplotypes at these polymorphic positions.

The 15µl PCRs contained 1µM of each primer, the standard PCR buffer (see chapter 2) and 0.1 U/µl of Taq polymerase. The reactions were cycled 32 times using a Geneamp 9600^{TM} thermal cycler (Perkin Elmer), each cycle consisted of denaturation at 96°C for 20 seconds, primer annealing at 67.5°C for 40 seconds and extension at 70°C for 50 seconds. The PCRs were split equally into four restriction enzyme digests using the appropriate buffer and conditions recommended by the manufacture. The products of the reactions were resolved by electrophoresis through a 2.2% metaphor agarose gel (FMC) in 0.5×TBE and the fragments observed by staining the gel with ethidium bromide.



Chapter 3. Figure 4

Table 3.4a. Haplotypes A, B and C in Caucasians including values of linkage disequilibrium (D-value)

Haplotype	-842 *DdeI*	-652 *AvaII*	-427 *TaqI*	-414 *MboII*	D-value -842 to -427	$\chi 2$	D-value -842 to -414	$\chi 2$
A	C	G	G	C	0.22	65	0.25	80
B	A	A	A	T	0.22	65	0.25	80
C	A	A	G	T	-0.22	65	0.25	80
$\chi 2$ test for over all hypothesis of linkage disequilibrium						65		80
						P< 0.01		P< 0.01

Table 3.4b. Observed and expected genotype frequencies of haplotypes A, B and C in Caucasians

Caucasians (n=80)		
Genotype	Observed	Expected
A/A	22	21
A/B	33	36
A/C	4	4
B/B	17	16
B/C	4	4
C/C	0	0
Total	80	
Frequencies		
A	0.51	
B	0.44	
C	0.05	
$\chi 2$=	0.71	P=0.98

account for the reduced efficiency of amplifications by TSK8B in some DNAs. An amplification refractory mutation system (ARMS) allele specific primer was synthesised (TS-13AR) to assay the -13 position, to provide a system for haplotype analysis, and to allow a reliable assay for the -176 polymorphism to be designed.

The -176 polymorphism. All Caucasian DNAs (n=80) that were homozygous for haplotypes A, B and C were homozygous at the -176 *DdeI* polymorphism. PCR amplification with the primer TS-13AR with TSK8C allows allele specific amplification from the SINE into the minisatellite in individuals heterozygous at -13 (Fig. 3.3.). This allowed haplotype analysis of the -176 position to be undertaken. Again, in all the heterozygous individuals analysed no new haplotypes were identified. The -176 polymorphism in Caucasians is in complete linkage disequilibrium with haplotypes A, B and C identified at the -842, -652, -427 and -414 polymorphic positions.

The -30 and -13 polymorphisms. Analysis of the -30 and -13 polymorphisms (Fig. 3.2.) using ARMS (Newton *et al.*, 1989) in Caucasians, revealed that both of these polymorphisms showed concordance with haplotypes A, B and C. All of the 79 haplotype B or C alleles had a T at these positions and of the 81 haplotype A alleles analysed, 77 had A at the -30 and -13 positions. The four haplotype A alleles that revealed additional heterogeneity consisted of three that had As at both -13 and -30 instead of Ts (designated haplotype A′), and one that had an A at the -13 position instead of T. This demonstrates that there is strong linkage disequilibrium in Caucasians across 7 polymorphic positions spanning 829bp (-13 to -842).

3. Analysis of the -842, -652, -427, -414 and -30 polymorphisms in the Japanese, West-African and Afro-Caribbean populations.

To observe whether the strong linkage disequilibrium observed in Caucasians was widespread, the analysis of these polymorphic positions was extended to other populations.

The -842, -652, -427 and -414 polymorphisms. RFLP analysis in Japanese, West-Africans and Afro-Caribbeans at the -842, -652, -427 and -414 positions was carried out. These positions were polymorphic in these populations and most individuals that were homozygous at one position were homozygous at the other three; no new haplotypes were identified. The haplotypes were in Hardy-Weinberg equilibrium in all the populations ($p>0.05$)(Table 3.5.). The Afro-Caribbeans showed a small deviation from Hardy-Weinberg equilibrium but the small numbers in some classes may have unduly influenced the χ^2 value. After

Table 3.5. Genotype frequencies of Haplotypes A, B and C

Japanese (n=106)			West-Africans (n=27)			Afro-Caribbeans (n=38)		
Genotype	Observed	Expected	Genotype	Observed	Expected	Genotype	Observed	Expected
A/A	12	9	A/A	16	15	A/A	23	21
A/B	38	44	A/B	1	1	A/B	3	8
A/C	0	0	A/C	7	10	A/C	7	7
B/B	55	52	B/B	0	0	B/B	4	1
B/C	1	1	B/C	0	0	B/C	0	1
C/C	0	0	C/C	3	2	C/C	1	1
Total	106		Total	27		Total	38	
Frequencies			Frequencies			Frequencies		
A	0.29		A	0.74		A	0.74	
B	0.70		B	0.02		B	0.14	
C	0.00		C	0.24		C	0.12	
χ^2=	2.22	P=0.82	χ^2=	2.47	P=0.78	χ^2=	3.72	P=0.16

pooling small classes with a value of less than five, a non-significant deviation from Hardy-Weinberg equilibrium was found. The Afro-Caribbean DNA samples were obtained from individuals involved in paternity cases in the London area, these individuals were identified as Afro-Caribbean from photographs. It is possible that these racial identifications were not wholly accurate and therefore these samples may not be a true representation of the Afro-Caribbean population. Therefore, admixture could explain why this population did not appear to conform as well as that expected from a population in Hardy-Weinberg equilibrium.

RFLP analysis on positions -842, -652, -427 and -414 was also carried out on DNA from the Amazonian Karitiana population. The Karitiana have a much reduced heterozygosity at other nuclear polymorphic loci compared to most human populations (Kidd *et al.*, 1991; Callegarijacques *et al.*, 1994). It is presumed that the Karitiana gene pool has been severely affected by founder effect and random drift resulting in a reduction of variation. Analysis of the DNA flanking the Xp:Yp telomere demonstrated that this population (n=22) was fixed for haplotype B (data not shown).

The -30 polymorphism. Analysis of the -30 position was also carried out on DNAs from Japanese and West-African individuals. In the 99 Japanese haplotype B alleles assayed all had an A at the -30 position; 47 of the 51 haplotype A alleles assayed had a T at the -30 position while four (8%) had an A at the -30 position (designated haplotype A´) (Table 3.6.). Analysis of the -30 position in West-Africans again showed that 100% of haplotype B and C alleles had an A at this position and demonstrated the existence of haplotypes A and A´. The haplotype A´ allele frequency was 0.24 (Table 3.6.).

To summarise, these analyses have demonstrated significant linkage disequilibrium exists across seven polymorphic positions (-842, -652, -427, -414, -176, -30 and -13) (χ^2=74.22, p<0.01) spanning 829bp immediately adjacent to the Xp:Yp pseudoautosomal telomere in Caucasians (Table 3.6.). In the Japanese, significant linkage disequilibrium has been demonstrated across five polymorphic sites (-842, -652, -427, -414 and -30) spanning 812bp (χ^2=65.42, p<0.01) (Table 3.6.), while the linkage disequilibrium across the same five polymorphic positions was reduced in the West-African population (χ^2=9.45, p<0.1)(Table 3.6.).

Table 3.6. Allele frequencies of haplotypes A, A', B and C in three different populations including values of linkage disequilibrium (D)

Haplotype	-842 *DdeI*	-652 *AvaII*	-427 *TaqI*	-414 *MboII*	-30 ARMS	Caucasians (n=80) Frequency	D value -842 to -30	$\chi 2$	Japanese (n=76) Frequency	D value -842 to -30	$\chi 2$	West-Africans (n=27) Frequency	D value -842 to -30	$\chi 2$
A	C	G	G	C	T	0.49	0.24	74.22	0.31	0.21	65.42	0.50	0.13	9.45
A'	C	G	G	C	A	0.02	-0.24	74.22	0.03	-0.21	65.42	0.24	-0.13	9.45
B	A	A	A	T	A	0.44	0.24	74.22	0.66	0.21	65.42	0.02	0.13	9.45
C	A	A	G	T	A	0.05	0.24	74.22	0.00	0.21	65.42	0.24	0.13	9.45
$\chi 2$ test for over all hypothesis of linkage disequilibrium								74.22			65.42			9.45
							P<	0.01		P<	0.01		P<	0.1

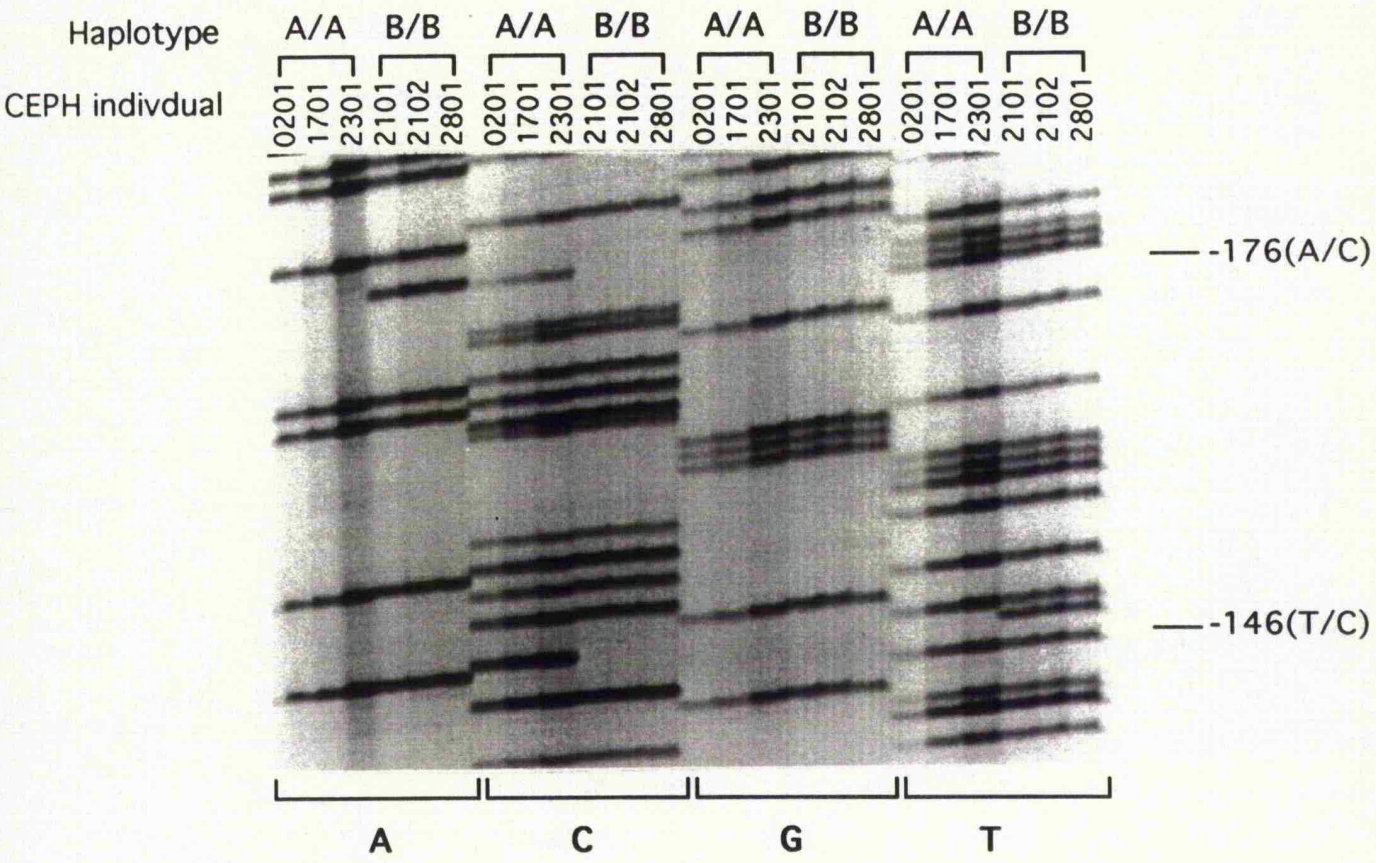
4. Sequence Divergence between Haplotypes

Sequence Analysis. To determine whether the other polymorphisms identified in the DNA adjacent to the Xp:Yp telomere also fell into just three haplotypes, sequence analysis was carried out. Three Caucasian DNAs homozygous for haplotype A (6 alleles) and three homozygous for haplotype B (6 alleles) were sequenced across the minisatellite (primers TSK8C and TSK8G) and the SINE (primers TSK8E and TSK8B)(Fig. 3.2b.). Haplotype C, alleles which have a low frequency in Caucasians, were sequenced by allele specific amplification from three individuals heterozygous for AC, using primer TS-13AR. Remarkably, all the alleles of any one haplotype had the same sequence at all the intervening polymorphic positions (Fig. 3.5.)(Table 3.7.). It is therefore possible that the 13 polymorphic positions in Caucasians are also in complete linkage disequilibrium resulting in just three highly diverged haplotypes. The degree of sequence divergence can be calculated between the three Caucasian alleles across 846bp (0-846) from the start of the telomere. Haplotypes A and B differ at 13 positions with a divergence of 1.5%, A and C differ at eight positions with a divergence of 0.95% and B and C differ at five positions with a divergence of 0.59%.

Sequence analysis on the West-African haplotypes was also carried out. Three DNAs homozygous for haplotype A (6 alleles) and three DNAs homozygous for haplotype C (6 alleles) were sequenced. The haplotype B is rare in West-Africans so just one allele was sequenced by allele specific amplification in a DNA heterozygous for AB. Three Afro-Caribbean DNAs homozygous for haplotype B were also sequenced (total of 7 alleles). The sequence at the intervening polymorphisms was mostly the same within haplotypic groups. Additional heterogeneity was detected at the -30 polymorphism (detected previously by ARMS analysis for the -30 position giving rise to the haplotype A′) (Table 3.6.) and at the -544 and -540 positions (Table 3.7.). One DNA homozygous for haplotype A and one homozygous for haplotype C were heterozygous at -544 and -540. Additional sequence information was available from an individual homozygous for haplotype A from the initial screen for polymorphisms. This individual had G and T at the -544 and -540 positions respectively instead of C and C. This additional heterogeneity forms two more haplotypes in the West-African population, the frequency of these haplotypes has not been quantified. These additional haplotypes could have arisen as an inter-allelic conversion like event between the minisatellites in the West-African A and C haplotypes. It is possible to determine the maximum divergence between different West-African haplotypes. The divergence of a haplotype A allele with G and T at the -544 and -540 positions respectively from a haplotype B allele includes 17 single base changes and one

Fig. 3.5. Showing sequence analysis of the -176 and -146 polymorphisms flanking the Xp:Yp telomere.

Unrelated Caucasian DNAs homozygous for the haplotypes A and B were PCR amplified with primers TSK8E and TSK8B. The products were purified by electroelution onto a dialysis membrane and sequenced with primer TSK8B (see chapter 2). The termination reactions for a single base from each sample were loaded in adjacent lanes to improve the detection polymorphisms. The bottom strand is shown in this photograph. Two polymorphisms are shown; all individuals homozygous for haplotype A had a C (G top strand) at -146 and a C (G top strand) at -176; all individuals homozygous for haplotype B had a T (A top strand) at -146 and an A (T top strand) at -176.



Chapter 3. Figure 5

Table 3.7.

							<--TSK8G TSK8E-->												
position from first telomere repeat	-842 *DdeI* +/-	-652 *Ava II* +/-	-554	-544	-540	-427 *TaqI* +/-	-415 *Mbo II* +/-	Δ 373-363[a]	-338	-333	-298	-297	-176 *DdeI* +/-	-147	-146	-145	-73	-30 ARMS	-13 ARMS
Nucleotides at the polymorphic positions 5' -> 3' -> telomere.																			
Caucasian	C/A	G/A	A/T	G/C	T/C	G/A	C/T	+	G	A	A/G	T/C	G/T	A	G/A	T	C	T/A	T/A
African	C/A	G/A	A/T	G/C	T/C	G/A	C/T	+/-	A/G	C/A	A/G	T/C	G/T	G/A	G/A	C/T	G/C	T/A	NT
Haplotypes from sequence analysis.																			
Caucasian Haplotypes																			
A	C	G	A	G	T	G	C	+	G	A	A	T	G	A	G	T	C	T	T
B	A	A	T	C	C	A	T	+	G	A	G	C	T	A	A	T	C	A	A
C	A	A	A	G	T	G	T	+	G	A	G	C	T	A	G	T	C	A	A
West-African Haplotypes																			
A	C	G	A	C G	C T	G	C	-	A	C	A	T	G	G	G	C	G	T A	NT
B	A	A	T	C	C	A	T	+	G	A	G	C	T	A	A	T	C	A	NT
C	A	A	A	G C	T C	G	T	+	G	A	G	C	T	A	G	T	C	A	NT

[a] The 10 bases between -363 and -373 are present (+) or absent (-) as indicated.

NT = not tested

insertion/deletion of 10bp in the 820bp proximal to the telomere (from -26 to -846). This represents a sequence divergence of 2.2%.

The Caucasian and West-African B haplotypes were identical across the 820bp sequenced as were the Caucasian and West-African C haplotypes (Table 3.7.). However, sequence comparison of haplotype A revealed differences between the two populations. Excluding the -544 and -540 positions, there are six differences arising from positions that are polymorphic in West-Africans but monomorphic in Caucasians. This represents a sequence divergence of 0.73%.

During the course of the sequence analysis additional variant positions were identified. One at position -826 (C-T) occurred only in West-African haplotype A alleles with a frequency of 0.36 (n=18). Another at -217 (G-C) observed only in Caucasian haplotype A alleles with a frequency of 0.073 (n=41), the West-Africans were not typed for this polymorphism. Other rare variants were identified during the course of the RFLP analysis which changed restriction sites but each was only detected once.

A high frequency of sequence polymorphism has been detected in the 846bp immediately flanking the telomere. This, coupled with significant linkage disequilibrium, results in a limited number of highly diverged haplotypes in each population examined

5. Defining the proximal limit of the high frequency of base substitutional polymorphisms and linkage disequilibria.

Primers TSK8J and TSK8K (2.0kb and 1.5kb proximal to the telomere respectively) were synthesised using sequence information from the clone T7A1/4 (Brown *et al.*, 1990). Sequence analysis of 320bp between these primers, from TSK8J was carried out on DNA amplified from 17 Caucasian and 11 West-African DNAs and two polymorphisms were identified. The -1888 (G-A) polymorphism, was present in both populations and the -1987 (G-A) polymorphism was only identified in West-Africans. The frequency of polymorphisms (1 in 320bp in Caucasians; 1 in 160bp in West-Africans) is reduced compared to that observed in the 846bp flanking the telomere. However, the frequency of polymorphisms though reduced, are not significantly different from the frequencies in the 846bp flanking the telomere (Caucasians p=0.224 and for West-Africans p=0.176 using a Fisher exact test; two tailed).

The -1888 polymorphism. The -1888 (G-A) polymorphism creates a *NlaIV* RFLP which can be assayed by PCR amplification with primers TSK8J and TSK8K

followed by digestion with *NlaIV* (Fig. 3.6.). Analysis of the -1888 RFLP in Caucasians, West-Africans, Afro-Caribbeans and Japanese revealed that it was in Hardy-Weinberg equilibrium in each population ($p>0.05$)(Table 3.8.). Haplotype analysis was carried out in Caucasians by allele specific PCR amplification of individuals doubly heterozygous for the -1888 and the polymorphisms in the 847bp flanking the telomere. The haplotype analysis was performed in Caucasians using primers TS-13AR and TSK8J followed by digestion with *NlaIV* (Fig. 3.6.). This analysis allowed linkage disequilibrium values to be calculated across the entire region from the -13 to the -1888 polymorphism. In Caucasians significant linkage disequilibrium between the -13 and -1888 positions was observed ($\chi^2=39.21$, $p<0.01$)(Table 3.9.). Haplotype analysis was not carried out directly in the Japanese population but it was possible to estimate linkage disequilibrium values between the -30 and -1888 polymorphisms, after making the assumption that all doubly heterozygous individuals (A/B and +/-) had haplotypes A+ and B-. This assumption seems valid because all of the Japanese individuals (n=31) homozygous for haplotype B were homozygous for the -1888 *NlaIV*- allele. Significant linkage disequilibrium between the -30 to the -1888 polymorphism was detected in the Japanese population ($\chi^2=25.53$, $p<0.05$)(Table 3.10.). However, comparison of the frequencies of the flanking haplotypes A, B and C with the frequencies of the + and - alleles at the -1888 polymorphism in West-Africans and Afro-Caribbeans (Tables 3.5. and 3.8.), suggests that the level of linkage disequilibrium is reduced in these populations compared to the Caucasian and Japanese populations.

6. Segregation analysis of DXYS14 and the Xp:Yp telomere flanking haplotypes.

In order to examine meiotic recombination rates in the terminal 13.4kb of the Xp:Yp pseudoautosomal region, the segregation of the haplotypes identified adjacent to the Xp:Yp telomere was compared to that of DXYS14 alleles in pedigrees of the CEPH panel of DNAs. Segregation data of alleles at DXYS14, that had been contributed to the CEPH database version 60, were used. Pedigrees informative for the telomere adjacent haplotypes were selected and typed for the RFLPs that distinguish the haplotypes A, B and C and compared to the segregation of DXYS14 alleles. This demonstrated that haplotypes A, B and C segregate through pedigrees in a Mendelian fashion. A total of 331 meioses were examined and no recombinants were identified, therefore, it is not possible to make a meaningful estimate of recombination rates between these two loci. The interval between DXYS14 and DXYS78, the next most poximal locus on the genetic map, is approximately 57-67kb. Two recombinants (both paternal) have been identified between these two loci (Vergnaud, 1994). Given the short physical distance

Fig. 3.6. The RFLP assay and haplotype analysis for the -1888 polymorphism flanking the Xp:Yp telomere.

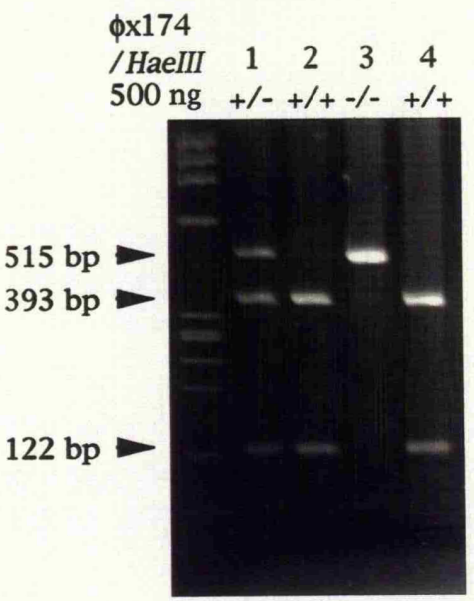
A. The -1888 *NlaIV* RFLP assay. DNAs were amplified with primers TSK8K and TSK8J and digested with *NlaIV*. The presence of the *NlaIV* site at the -1888 position cuts the 515 bp PCR product into a 393 bp fragment and a 122 bp fragment, the analysis of four DNAs is shown in A.

The 10µl PCRs contained 1µM of each primer, the standard PCR buffer (see chapter 2) and 0.1 U/µl of Taq polymerase. The reactions were cycled 32 times using a Geneamp 9600^{TM} thermal cycler (Perkin Elmer), each cycle consisted of denaturation at 96°C for 20 seconds, primer annealing at 68°C for 40 seconds and extension at 70°C for 60 seconds. 5µls of the PCR products were digested with *NlaIV* using the appropriate buffer and conditions recommended by the manufacture. The products of the reactions were resolved by electrophoresis through a 1.9% metaphor agarose gel (FMC) in 0.5×TBE and the fragments observed by staining the gel with ethidium bromide.

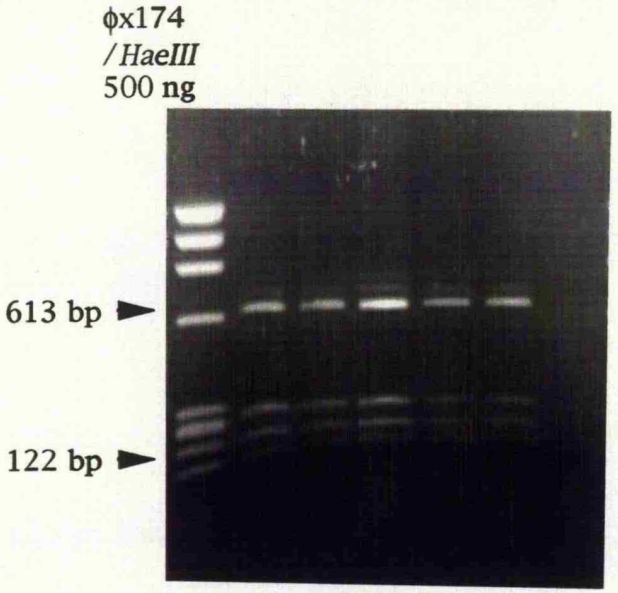
B. The haplotype analysis of the -1888 polymorphism was achieved by allele-specific amplification of individuals heterozygous at the -13 polymorphic position with primers TS-13AR and TSK8J followed by digestion with *NlaIV*. The analysis of five DNAs is shown in B. The 1.9 kb PCR product is cut multiple times with *NlaIV* the presence of an *NlaIV* site at the -1888 position creates two products of 613 bp and 122 bp and the other digestion products were monomorphic.

The 10µl PCRs contained 1µM of each primer, the standard PCR buffer (see chapter 2) and 0.1 U/µl of Taq polymerase. The reactions were cycled 32 times using a Geneamp 9600^{TM} thermal cycler (Perkin Elmer), each cycle consisted of denaturation at 96°C for 20 seconds, primer annealing at 67°C for 40 seconds and extension at 70°C for 2 minutes 20 seconds. 5µls of the PCR products were digested with *NlaIV* using the appropriate buffer and conditions recommended by the manufacture. The products of the reactions were resolved by electrophoresis through a 2% metaphor agarose gel (FMC) in 0.5×TBE and the fragments observed by staining the gel with ethidium bromide.

A. -1888 *NlaIV* +/-



B. -1888 *NlaIV* haplotype analysis



Chapter 3. Figure 6

Table 3.8. Genotype frequencies at the -1888 polymorphism in Caucasians, Japanese, West-Africans and Afro-Caribbeans.

Caucasians (n=80)			Japanese (n=103)		
Genotype	Observed	Expected	Genotype	Observed	Expected
+/+	30	26	+/+	7	5
+/-	31	39	+/-	32	36
-/-	19	15	-/-	64	62
Total	80		Total	103	
Frequencies			Frequencies		
+	0.57		+	0.22	
-	0.43		-	0.78	
$\chi 2=$	3.53 P=	0.17	$\chi 2=$	1.12 P=	0.57

West-Africans (n=27)			Afro-Caribbeans (n=37)		
Genotype	Observed	Expected	Genotype	Observed	Expected
+/+	3	3	+/+	10	6
+/-	13	12	+/-	11	18
-/-	11	11	-/-	16	12
Total	27		Total	37	
Frequencies			Frequencies		
+	0.35		+	0.42	
-	0.65		-	0.58	
$\chi 2=$	0.08 P=	0.96	$\chi 2=$	5.61 P=	0.06

Table 3.9. Linkage disequilibrium between telomere flanking haplotypes and the -1888 polymorphsim in Caucasians.

-842, -652, -427 -414, -30, -13 Haplotype	-1888 *NLAIV*	Observed frequency	Observed number	Expected number	D value	$\chi 2$
A	+	0.46	71	43	0.174	38.35
A	-	0.04	6	33	-0.174	38.35
A'	+	0.01	1	2	-0.005	0.35
A'	-	0.01	2	1	0.005	0.35
B	+	0.11	17	39	-0.143	26.53
B	-	0.33	52	30	0.143	26.53
C	+	0.00	0	4	-0.026	4.87
C	-	0.04	7	3	0.026	4.87
$\chi 2$ test for over all hypothesis of linkage disequilibrium (d.f.=3)						39.21
					P<	0.01

Table 3.10. Linkage disequilibrium between telomere flanking haplotypes and the -1888 polymorphsim in Japanese.

-842, -652, -427 -414, -30 Haplotype	-1888 *NLAIV*	Observed frequency	Observed number	Expected number	D value	$\chi 2$
A	+	0.23	14	4	0.162	20.64
A	-	0.03	2	12	-0.162	20.64
A'	+	0.03	2	1	0.024	2.84
A'	-	0.00	0	1	-0.024	2.84
B	+	0.00	0	11	-0.187	25.45
B	-	0.70	42	31	0.187	25.45
$\chi 2$ test for over all hypothesis of linkage disequilibrium (d.f.=2)						25.53
					P<	0.05

(13.4kb) between DXYS14 and the telomere it is not surprising that no recombination events were identified. However, it appears that there may not be an increase in recombination rates between DXYS14 and the telomere compared to that between DXYS14 and DXYS78.

7. Similarity to other sequences.

An investigation as to whether the sequences flanking the Xp:Yp pseudoautosomal telomere have homology to other sequenced regions of the human genome was carried out. The sequence of the clone T7A1/4 (Brown *et al.*, 1990) was submitted for a sequence similarity blast search (Altschul *et al.*, 1990) at the National Centre for Biotechnology Information (Bethesda, MD) through the Brookhaven Protein data bank, Genbank (release 90) and EMBL data library (release 43). As expected, the region covered by the SINE of which there are an estimated 20,000 copies (La Mantia *et al.*, 1989) (Fig. 3.1.), showed similarity to multiple transcribed sequences. The proximal 202bp of the T7A1/4 sequence showed an 88% sequence similarity to the cDNA clone yi54b04.s1 (Hillier *et al.*, 1995). This clone had an additional 253bp which extended proximal to the first base of the T7A1/4 sequence. An additional search for similarity through the EST database revealed a further four cDNA clones all showing a similarity of 87-88 % with the most proximal 200bp of the T7A1/4 sequence. All of these cDNA clones extended proximal to the first base of the T7A1/4 sequence. It appears that the most proximal 200bp of the T7A1/4 sequence has similarity with a family of transcribed sequences.

Discussion

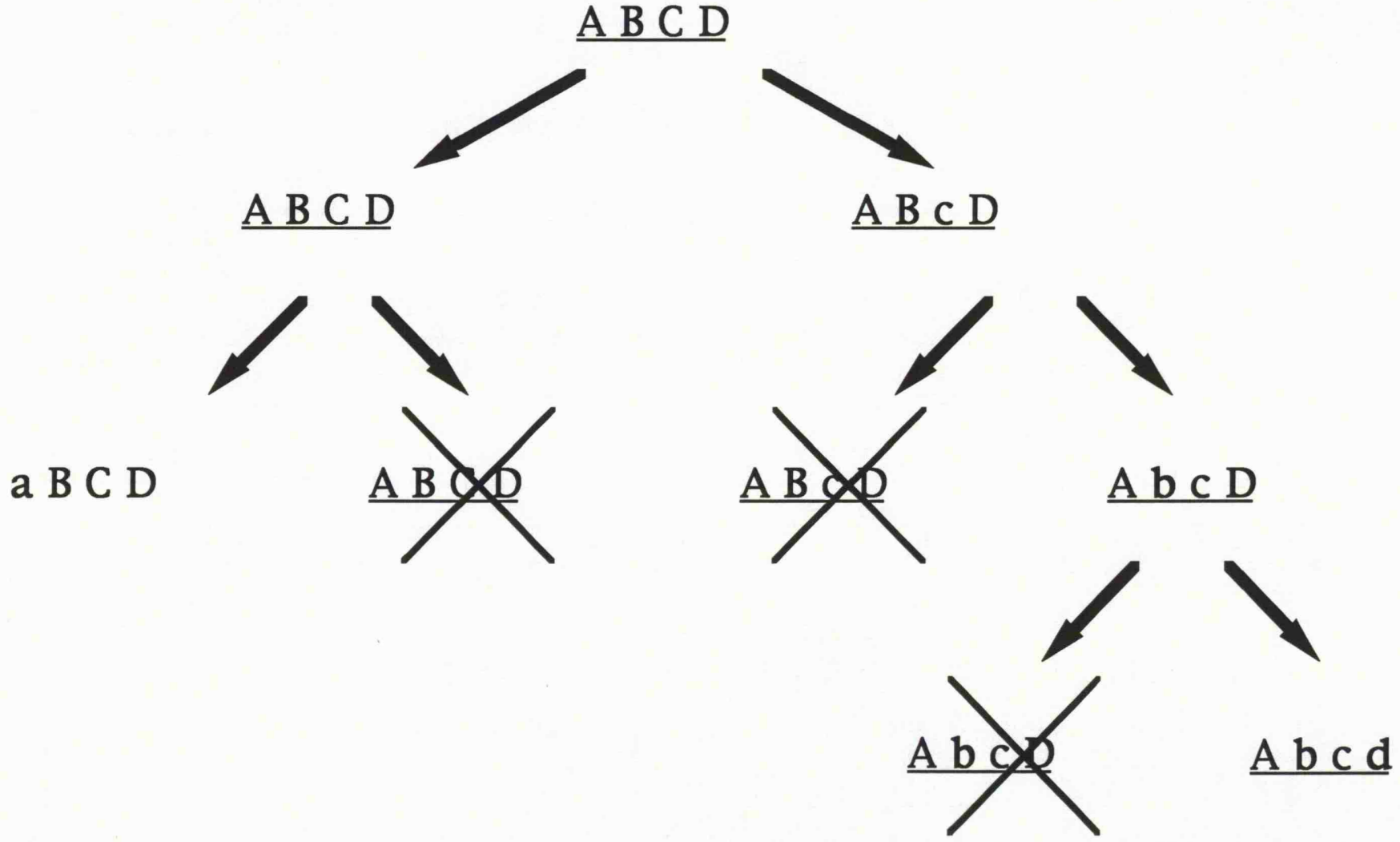
Sequence polymorphism. A high frequency of sequence polymorphism in the DNA immediately flanking the Xp:Yp pseudoautosomal telomere has been identified. Estimates of the frequency of sequence polymorphism in other regions of the human genome vary from one in 200bp (Miyamoto *et al.*, 1988) to one in 1000bp (Cooper *et al.*, 1985). There are exceptions, sequence polymorphism at the HLA loci is much greater than that detected elsewhere in the human genome, with a frequency of approximately one polymorphism every five bases over 190bp of coding sequence (Gyllensten and Erlich, 1988). These major histocompatibility complex (MHC) classic class 1 antigen loci encode glycoproteins play a central role in governing the immune system (Bjorkman *et al.*, 1987). A particular class 1 antigen will provide enhanced recognition for a particular antigen, individuals with different types of antigen will have a selective advantage (Enssle *et al.*, 1987). This could account for the maintenance of polymorphism at these loci by providing a basis for balancing selection to operate such as overdominant selection (Nei, 1988) or frequency dependent selection (Bodmer, 1972; Slade and McCallum, 1992). How the high frequency of sequence polymorphism at the Xp:Yp telomere junction arose and is maintained is unclear. There is no mechanistic basis to explain the high frequency of polymorphism in a similar manner to that observed at the HLA loci. It is possible that the high frequency of polymorphisms in DNA flanking the Xp:Yp telomere could be due to a reduction in the availability of this DNA to the mechanisms involved in DNA repair. This could be a consequence of its position at the end of the chromosome, which in other eukaryotes exhibit heterochromatic characteristics such as condensed chomatin structure (John *et al.*, 1985; Traverse and Pardue, 1989) and in yeast have proteins bound which might prevent access by other proteins (Gottschling, 1992). It is known that the telomeres of humans have an unusual chromatin structure towards the terminus, but with a more canonical structure in the proximal regions with closely spaced nucleosomes similar to that identified in rat (Makarov *et al.*, 1993; Tommerup *et al.*, 1994) and that human telomere repeat arrays are bound by a telomere binding protein (Zhong *et al.*, 1992; Chong *et al.*, 1995). It is not clear whether the high frequency of sequence polymorphism is a general feature of chromosome ends or is a feature unique to the Xp:Yp pseudoautosomal chromosome end. This is difficult to evaluate because to date, no comparative studies have been carried out on other chromosome ends.

Sequence divergence. This analysis has demonstrated that a few highly diverged haplotypes exist immediately flanking the Xp:Yp pseudoautosomal telomere. These show the same divergence as that detected between non-coding

sequences in humans and chimpanzees (*Pan troglodytes*). Comparisons of orthologous sequences upstream and downstream of the φη-globin gene region between humans and chimpanzees demonstrates a sequence divergence of 1.61% (Miyamoto *et al.*, 1988). The sequence divergence between Caucasian haplotypes A and B is 1.5% and the maximum divergence between West-African haplotypes is 2.2%. Assuming evolution along haploid lineages, this implies that these haplotypes may be as ancient. Calculations based upon the divergence of the φη-globin gene region, place the separation of humans and chimpanzees between 4.7 and 7.1 Myr ago (Miyamoto *et al.*, 1988). Sequence divergence of the α_1-antitrypsin and β- and δ-globin genes between humans and new world monkeys place the divergence of chimpanzees from the human lineage at 6.4 ± 2.6 Myr (Sakoyama *et al.*, 1987). It is therefore possible that the divergence of the haplotypes A, B and C occurred at a similar time as the divergence of chimpanzees from the human lineage.

Linkage disequilibrium. RFLP analysis revealed almost complete linkage disequilibrium across four polymorphic sites in Caucasians, Japanese, West-Africans and Afro-Caribbeans (positions -842, -652, -427 and -414). Significant linkage disequilibrium was also detected when the -30 and -1888 polymorphic positions were included in the analysis of Caucasians and Japanese ($p<0.01$ and $p<0.05$ respectively). Sequence analysis of a limited number of alleles revealed that the other intervening polymorphic positions were also in linkage disequilibrium with the haplotypes identified by RFLP analysis, creating a limited number of highly diverged haplotypes. Linkage disequilibrium can result from selection, genetic drift, admixture or suppression of recombination (Ohta, 1982; Thomson and Klitz, 1987). No one of these effects can explain the strong linkage disequilibrium and sequence divergence observed at this locus. There could be reduced recombination because few haplotypic intermediates that could be the result of recombination have been identified. However, a reduction in recombination cannot explain the observed divergence between the different haplotypes. The existence of multiple haplotypes could result from a series of sequential mutations of existing haplotypes to form new haplotypes. In order to explain the existence of a few highly diverged haplotypes, the model has to account for the loss of the existing haplotype coupled with the fixation of the modified haplotype multiple times (Fig. 3.7.). This is unlikely, given random genetic drift, and if these alleles have been present in the population for millions of years then their continued existence cannot be explained by neutral theory (Nei, 1987). Therefore, it is possible that some form of selection mechanism may be operating to maintain these haplotypes. If selection is having an effect on this locus, one would anticipate that this locus is located in or near to a gene. This gene would be maintained by selection in a limited number of ancient allelic

Fig. 3.7. Showing the formation of multiple haplotypes by sequential mutation.



states to account for the sequence divergence. This locus would be either in the coding or regulatory regions of the gene, or near enough for a strong hitch-hiking effect to take place (Thomson, 1977). The DXYS14 locus only 13.4kb from the Xp:Yp telomere has some sequence characteristics of a CpG island (Inglehearn and Cooke, 1990). However, transcripts that hybridise to DXYS14 were not detected in northern blot analysis of HeLa cell RNA and in DNA derived from blood the sequence is methylated, suggesting that DXYS14 is not a true CpG island (Inglehearn and Cooke, 1990). It may be possible that DXYS14 is hypomethylated in other tissues, such as the germ line. The possibility that the evolution of the DNA flanking the Xp:Yp telomere is affected by selection operating on a gene, is enhanced by the fact that the proximal 200bp of the clone T7A1/4 has similarity to multiple cDNA clones. Strong linkage disequilibrium exists from the -13 to the -1888 polymorphism in Caucasians and Japanese but the linkage disequilibrium seems to be significantly reduced in the West-African and Afro-Caribbean populations. This would argue against the existence of a gene under selection causing this effect, unless the -1888 position was located within an intron.

Allelic diversity at the HLA DQα locus appears to predate the speciation of the human, chimpanzee and gorilla lineages. The diversity appears to have been maintained, not by allelic recombination generating new variants, but by some selection mechanism such as overdominance (Hughes and Nei, 1988; Nei, 1988; Gyllensten and Erlich, 1989) or frequency dependent selection (Bodmer, 1972; Slade and McCallum, 1992). It is possible that haplotypes flanking the Xp:Yp telomere are also as old as the divergence of the chimpanzee, gorilla and human lineages, therefore, the selection mechanisms that maintain variation at the HLA DQα locus could be similar to those that maintain variation at the Xp:Yp pseudoautosomal telomere-junction.

The limited number of highly diverged haplotypes could be explained by a recent conversion event from another locus fixed for one of the haplotypes to the Xp:Yp pseudoautosomal telomere junction region fixed for a different haplotype. These haplotypes would be maintained by a suppression of recombination and their continual existence in the populations analysed could be explained with the above selection model. However, there is no evidence of duplication of the Xp:Yp pseudoautosomal telomere junction region in humans and therefore the hypothetical duplicated region would have to have been lost after the conversion event. Also, the haplotype A sequences obtained from Caucasians and West-Africans are different and more than two haplotypes have been identified, both these observations are not consistent with a conservation event from another locus.

The strong linkage disequilibrium observed at the Xp:Yp pseudoautosomal telomere junction region is somewhat paradoxical since there is a large body of evidence to suggest that the ends of human chromosomes are proficient at recombination (Hulten, 1974; NIH/CEPH, 1992; Litt *et al.*, 1995). However, these data do not include the most terminal sequences. To date, the ends of most genetic maps in humans do not include the telomeres and therefore no detailed knowledge on recombination rates upto a human telomere is available. The strong linkage disequilibrium and sequence divergence observed at the Xp:Yp pseudoautosomal telomere junction region, suggests that human telomeres may have a reduced rate of recombination compared to the proterminal regions. In addition, the Xp:Yp pseudoautosomal region has a gradient of recombination increasing toward the telomere, therefore, one may anticipate that the very end of the chromosome may have the highest recombination rate. The segregation data between DXYS14 and the telomere, suggest that there is no increase in recombination in this interval. However, it is not possible to say if there is a reduction in recombination in this interval.

Conclusions

A high frequency of sequence polymorphism has been detected in the 846bp of DNA immediately flanking the Xp:Yp pseudoautosomal telomere. These polymorphisms are in strong linkage disequilibrium, resulting in a few highly diverged haplotypes. The strong linkage disequilibrium extends to the proximal -1888 polymorphism in Caucasians and Japanese, but appears to be weaker in the Afro-Caribbean and West-African populations.

To further analyse variation in this region, detailed analysis of sequence polymorphism in the telomere itself was carried out. The development of a technique to access this variation is described in the next chapter.

Chapter 4

THE DEVELOPMENT OF TELOMERE VARIANT REPEAT MAPPING BY PCR (TVR-PCR)

Summary

Human telomeres are composed largely of the sequence TTAGGG tandemly repeated to form arrays between 5 to 20kb. It has been demonstrated that in addition to TTAGGG repeats, variant repeats occur but are localised in the proximal 2kb of telomere repeat arrays. In this chapter I have described the development of a technique to assay the distribution of TTAGGG, TGAGGG and TCAGGG repeats in the proximal 720bp of the Xp:Yp pseudoautosomal telomere repeat array. The products observed by TVR-PCR can be coded to form a databank of allele codes. The distribution of the telomere and variant repeats is shown to be highly variable. This variation raises the possibility of its use as a technique for individual identification for use in forensic and parentage analysis. The TVR-PCR technique will provide an informative system to observe the dynamics of human telomeres in the human population, as well as in somatic tissues and in the germline. TVR-PCR is in principal applicable to any human telomere and could potentially be used to analyse the telomeres of other organisms.

Introduction

The human telomerase catalyses the addition of TTAGGG repeats to the terminus of the chromosomes (Moyzis *et al.*, 1988; Morin, 1989; Feng *et al.*, 1995). However, human telomere arrays have been shown to be composed of additional variant repeats such as TGAGGG and TTGGGG. The variant repeats appear to be clustered in the proximal 2kb of human telomere repeat arrays (Allshire *et al.*, 1989). Analysis of human telomeres has revealed additional sequence heterogeneity, but the general features of a 6bp periodicity and the presence of a run of three G-residues are nearly always present (Brown *et al.*, 1990)(N. J. Royle pers. comm.). These features may reflect some functional constraints, for example, disruption in the periodicity of the G-residues may change the higher order structure of the telomere, and in turn alter the nucleosome and chromatin structure (Makarov *et al.*, 1993; Tommerup *et al.*, 1994). The repeat array may also be constrained by the specificity of telomere binding proteins, the mammalian telomere binding protein TRF will not bind repeats other than TTAGGG, including TTGGGG a common human telomere repeat variant (Zhong *et al.*, 1992; Chong et al., 1995). The introduction of a modified repeat sequence into the telomeres of the ciliate *Tetrahymena,* by the mutation of the RNA component of telomerase, resulted in impaired cell division and a senescence phenotype (Yu *et al.*, 1990). However, the presence of variant repeats at internal regions of *Tetrahymena* telomere repeat arrays has no effect on telomere function (Yu and Blackburn, 1991). The function of any of these variant telomere repeats is unknown, it has been suggested that they may have a role in the pairing of homologues and recombination during meiosis (Allshire *et al.*, 1989). How these repeats arose is also unclear. It is unlikely that they were erroneously catalysed by a human telomerase, which can faithfully catalyse the addition of TTAGGG repeats *in vitro* (Morin, 1989; Feng *et al.*, 1995). The variant repeats may have arisen by point mutations followed by expansion via intra-allelic mechanisms such as slippage during replication (Levinson and Gutman, 1987; Murphy *et al.*, 1989; Schlotterer and Tautz, 1992) and unequal sister-chromatid exchange (Schlotterer and Tautz, 1994), or inter-allelic mechanisms such as unequal-homologous exchanges and gene-conversion like events (Carpenter, 1987; Jeffreys *et al.*, 1994). If these repeats are present at the end of all chromosomes then this could reflect recombinogenic interactions between chromosome ends, which has been demonstrated in yeast (Pluta and Zakian, 1989; Louis and Haber, 1990; Lundblad and Blackburn, 1993) and may occur in humans (Wilkie *et al.*, 1991; Ijdo *et al.*, 1992). This possibility is further backed up by the presence of subterminal repeat families present at multiple chromosome ends (Brown *et al.*, 1990; de Lange *et al.*, 1990; Ijdo *et al.*, 1992; Royle *et al.*, 1992). However, sequence divergence of

upto 20% exists between subterminal repeats. This suggests that only a low level of homogenisation by exchange between nonhomologous chromosome ends occurs and that these sequences must be old (Royle *et al.*, 1992). The telomeres of some mice strains can be very long, upto 150kb (Kipling and Cooke, 1990). The long telomeres can be resolved as single bands by PFGE, the bands are hypervariable such that inbred mouse strains are heterozygous at all the telomeres analysed by this technique (Kipling and Cooke, 1990; Starling *et al.*, 1990). The underlying cause of this hypervariability must be a high mutation rate often involving large changes in size (Starling *et al.*, 1990). The telomeres of mice can be cleaved with the restriction enzyme *Mnl1* which has the recognition site GAGG and therefore it will only cut the mouse telomere sequence if there is a variant repeat. Cleavage of mouse telomeres with this enzyme reveals hypervariable fragments which may be generated by mechanisms similar to those that generate the variability in the full length telomeres (Starling *et al.*, 1990). Speculation as to the mechanisms involved include inter and intra-allelic recombination and gene-conversion processes between telomeres, replication slippage and reduction of the telomere followed by expansion by telomerase (Starling *et al.*, 1990).

Mutation at short tandem repeat loci in humans. Repeat sequence mutation has recently been discovered to be a common cause of human genetic disease. The majority of these diseases involve instability at Short Tandem Repeat (STR) loci. This has resulted in intense interest in the mechanisms of repeat sequence turnover, particularly at trinucleotide repeat loci. Instability of trinucleotide repeat loci accounts for a growing number of diseases such as the Fragile X syndrome (Yu *et al.*, 1991) and Myotonic Dystrophy (Brook *et al.*, 1992). This instability has been referred to as dynamic mutation (Richards and Sutherland, 1992) since the probability of mutation is a function of repeat copy number. One possible mechanism for this is slippage during replication, the longer the repeat array the greater the probability that an Okazaki fragment will be situated within it, resulting in two single stranded breaks. The fragment is no longer anchored in unique sequence DNA and so is 'free' to slide during polymerisation. This results in a larger change in the repeat copy number than slippage in a smaller repeat array, where one part of the Okazaki fragment would be anchored by the flanking DNA (Richards and Sutherland, 1994).

Another class of human disease in which tandem repeat instability occurs has been observed in some tumour tissues, such as Hereditary Non-Polyposis Colon Cancer (HNPCC) (Aaltonen *et al.*, 1993). These cancers are thought to be associated with a susceptibility to DNA replication errors, such tumours have been termed RER+ (replication error repair). Such RER+ tumours exhibit instability in mono-,

di- and trinucleotide loci (Thibodeau *et al.*, 1993; Ionov *et al.*, 1993) but no instability has been observed at minisatellite loci (Allen, 1994) with the exception of the D1S7 locus which contains a 9bp repeat unit minisatellite (Hoffolsen *et al.*, 1995). It has been proposed that the mutation observed at STR loci in RER+ tumours is due to replication slippage events that are not repaired by cellular mismatch repair systems (Parsons *et al.*, 1993; Strand *et al.*, 1993; Shibata *et al.*, 1994).

Variation at minisatellite loci. Research to elucidate the mechanisms that generate hypervariability at minisatellite loci has been greatly facilitated by the development of a system to assay the distribution of different types of repeat unit within the minisatellite array. This system, called Minisatellite Variant Repeat mapping by PCR (MVR-PCR), has generated some remarkable insights into the mutation processes at several minisatellite loci (Jeffreys *et al.*, 1990, 1991; Armour *et al.*, 1993; Neil and Jeffreys, 1993). MVR-PCR revealed even greater variation than that previously detected by allele length analysis. The majority of the variation in repeat type distribution (and therefore mutation) was found to be restricted to one end of the minisatellite array at the loci studied. This implies that external elements may mediate minisatellite mutation. Further analysis of mutants isolated from pedigrees and sperm shows that the majority of mutations are complex inter-allelic gene-conversion-like events (Jeffreys *et al.*, 1994).

This work. In this chapter I have described the development of a technique similar to MVR-PCR to assay the distribution of different types of telomere repeat in the proximal regions of the Xp:Yp telomere repeat array. This assay is called Telomere Variant Repeat mapping by PCR (TVR-PCR). The possible applications of this technique are discussed.

Results

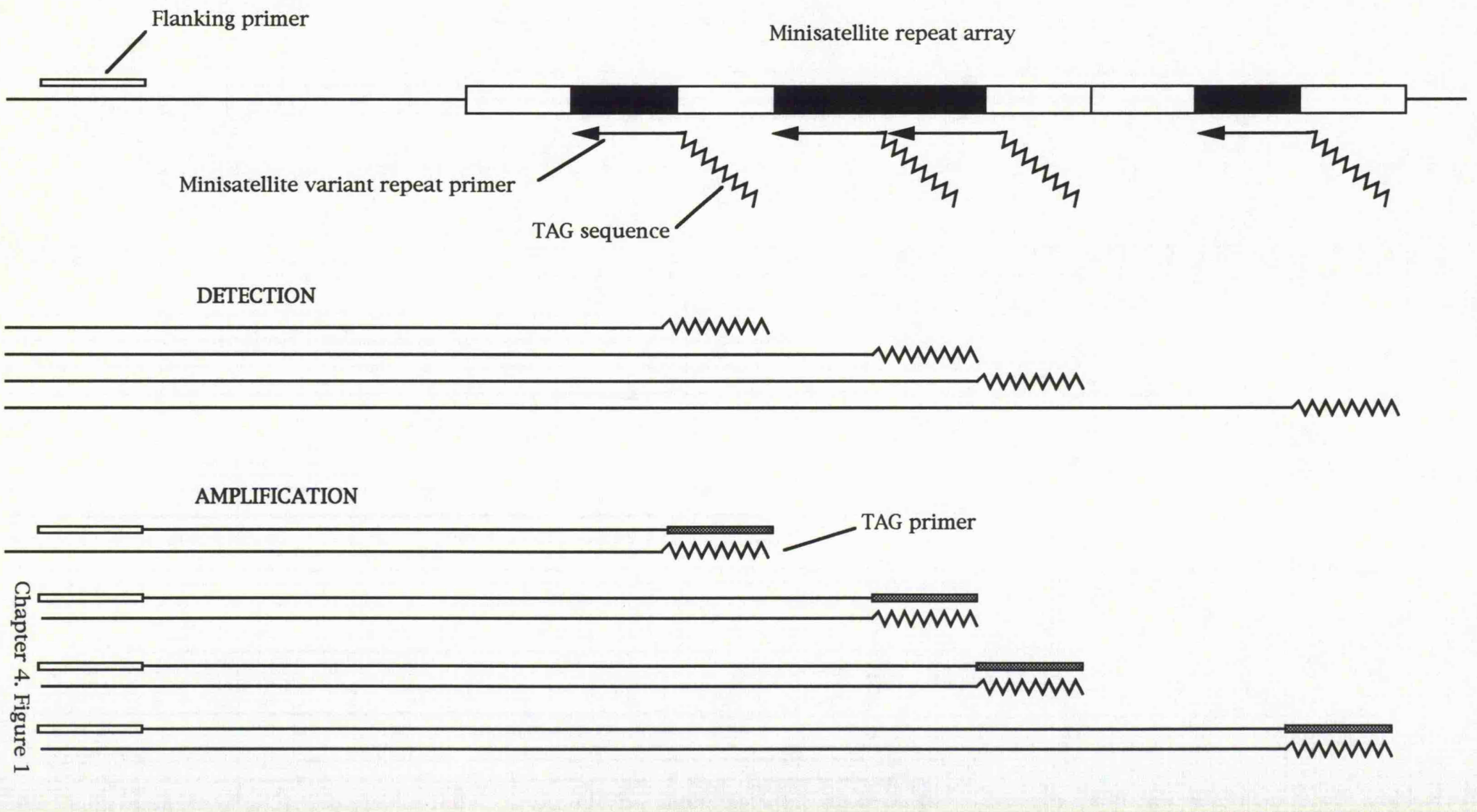
1. The principals of Minisatellite Variant Repeat mapping by PCR (MVR-PCR).

MVR-PCR is a technique used to assay the distribution of different types of minisatellite repeat unit (Jeffreys *et al.*, 1991). It is a PCR based technique using a locus specific primer in the DNA flanking the minisatellite repeat array. The primer is used in conjunction with primers designed to anneal to the different minisatellite repeat types, these primers have a 20 base 5′ extension called the TAG sequence. The minisatellite repeat primers are used at low concentrations (e.g. 10nM) with a flanking primer and a primer with the same sequence as the 5′ TAG sequence, both at a higher concentration (e.g. 1μM). The use of the 5′ TAG sequence in conjunction with the TAG primer uncouples detection of the repeat type from amplification and may prevent progressive shortening of the PCR products at each cycle, due to the repeat primer priming internally in the PCR products. The MVR-PCR products are separated by gel electrophoresis and detected by Southern blot hybridisation to a minisatellite repeat probe. A complementary ladder of products is generated from two or more PCRs, which corresponds to the position of the different repeat types in the minisatellite array (Fig. 4.1.). The pattern can be interpreted into a binary code which is amenable to databasing.

2. Telomere Variant Repeat mapping by PCR (TVR-PCR).

In order to develop a technique to assay the different types of repeat unit in the Xp:Yp telomere repeat array certain problems had to be overcome. The telomere repeat unit is just 6bp, therefore the repeat primers would span several repeat units, so careful primer design was needed so that single repeat units could be detected. The small repeat unit made the resolution of a set of products differing in size by 6bp difficult. At minisatellite loci single alleles could be analysed by MVR-PCR since they could be size separated and purified, this was not possible for the analysis of telomeric alleles. The telomere repeat primers initiate synthesis in the PCR from every telomere present in a diploid individual including interstitial telomere repeats, therefore, the specificity of the reaction relies entirely on the PCR primer in the flanking DNA. Additional locus specificity can only be obtained by Southern blot hybridisation using a probe containing flanking DNA, as a telomere repeat probe would hybridise to the products derived from all the other telomeres. The 392bp of DNA immediately flanking the Xp:Yp telomere contains a SINE estimated to be present in 20,000 copies in the human genome (La Mantia *et al.*, 1989) and just 11bp separate the SINE sequence from a

Fig. 4.1. The principles of minisatellite variant repeat mapping



four repeat unit minisatellite (see figure 3.4.). This meant that the design of a locus specific primer in the flanking DNA might be difficult.

3. Initial investigation to determine whether variation could be detected in the Xp:Yp telomere repeat array

Initially TVR-PCR was attempted with the primer TAG-TelD (Table 4.1), a primer designed to anneal to TGAGGG repeats and contains the 5′ TAG sequence as used in MVR-PCR primers (Jeffreys *et al.*, 1991). This was used in a nested PCR using primers TSK8C and TSK8E. The primer TSK8C is located in unique DNA approximately 850bp from the start of the telomere and TSK8E is positioned in the DNA between the proximal end of the SINE and distal end of the minisatellite (Fig. 3.1). This nested approach was taken to maximise the specificity of the reaction. Here, several rounds of PCR were carried out with primer TSK8C and TAG-TelD and the products from this were then used to seed a PCR reaction with TSK8E, TAG-TelD and TAG. The products were resolved by agarose gel electrophoresis followed by Southern blotting and hybridisation with a probe consisting of a PCR product generated from the clone TSK8 (Royle *et al.*, 1992) using primers TSK8A and TSK8B. This was carried out on DNA from two unrelated individuals, each had a different pattern of PCR products (Fig. 4.2a.). As expected the resolution was poor because the products generated are 400bp or more long and differed in size by only 6bp. A smear of PCR products was observed in the agarose gel stained with ethidium bromide after only a limited number of PCR cycles, this was due to the lack of specificity in these reactions. This lack of specificity was also demonstrated by hybridisation to the telomere repeat probe (Fig. 4.2b.), the majority of hybridised fragments were probably single stranded products derived from TGAGGG repeats in other telomeres. Subsequent experiments showed that the presence of the TAG sequence was not being used to drive the amplification reaction. The use of TelD with out the TAG sequence produced better amplification than that achieved with TAG-TelD and the TAG primer (data not shown), so the TAG primer and the TAG-telomere primer were no longer used.

4. The resolution of a 6bp telomere repeat ladder

In the initial experiments it was clear that the resolution of the PCR products needed improvement as the PCR products differed in size by only 6bp. Experiments were carried out using TSK8E and TelD with resolution on a 40 cm 1% NuSieve agarose/1% agarose (FMC) gel, followed by Southern blot hybridisation with a flanking probe. This resulted in an improvement in the resolution, such that the thick strongly hybridising bands observed in Figure 4.2a. were now resolved into discrete fragments with a 6bp periodicity (Fig. 4.2c.). However, the

Table 4.1. Telomere repeat primers.

TAG-TelC	**TCATGCGTCCATGGTCCGGA**CTAACCCTAACCCTAACCCTAA
TAG-TelD	**TCATGCGTCCATGGTCCGGA**CCTCACCCTCACCCTCACCCTCA
TelC	**GGCCATCGATGAATT**CTAACCCTAACCCTAACCCTAA
TelD	**GGCCATCGATGAATT**CCTCACCCTCACCCTCACCCTCA
TelG	CCCTCACCCTCACCCTCACCCTC
TelH	CCCTAACCCTAACCCTAACCCTA
TelF	CCCIIACCCIIACCC(CT)(AC)ACCCTC
TelL	ACCIIACCCIIACCCIIACCCTC
TelK	CCCTGACCCTGACCCTGACCTG
TelO	CCCTIACCCTIACCCTNACCCTA
TelP	CCCTIACCCTIACCCTNACCCTG
TelR	CCCTIACCCTIACCCTNACCCTC
TelT	CCCTIACCCTIACCCTNACCCTT
TelW	CCCTTACCCTTACCCTNACCCTA
TelX	CCCTTACCCTTACCCTNACCCTC
TelY	CCCTTACCCTTACCCTNACCCTG
TAG-TelJ	**TCATGCGTCCATGGTCCGGA**ACCCCAACCCCAACCCCAACCCC
TAG-TelW	**TCATGCGTCCATGGTCCGGA**CCCTTACCCTTACCCTNACCCTA
TAG-TelLX	**TCATGCGTCCATGGTCCGGA**CCCTTACCCTTACCCTNACCCTC
TAG-TelY	**TCATGCGTCCATGGTCCGGA**CCCTTACCCTTACCCTNACCCTG

Figure 4.2. TVR-PCR assaying the distribution of TGAGGG repeats in the Xp:Yp telomere from unrelated individuals.

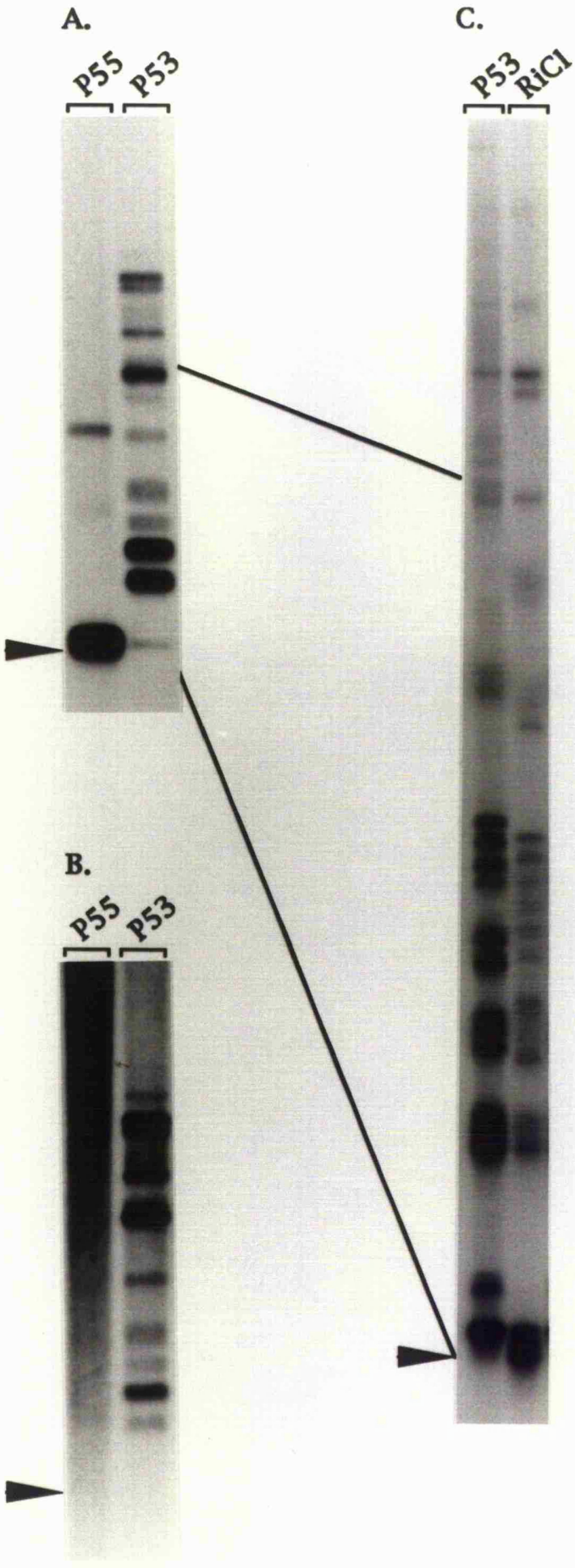
A. TVR-PCR with primer TAG-TelD designed to anneal to TGAGGG repeats, hybridised to a DNA probe containing the SINE. Two DNAs are shown both have different distributions of TGAGGG repeats. The large arrow on the left denotes the start of the telomere repeat array.

These reactions were carried out using a nested approach, with 5 cycles of PCR in the presence of flanking primer TSK8C and TAG-TelD and 12 cycles with the flanking primer TSK8E and TAG. The 7µl PCRs contained 0.1µM of the primers TSK8C and TAG-TelD, the standard PCR buffer (see chapter 2) and 0.1 U/µl of Taq polymerase. After the initial 5 cycles the reactions were made upto 10µl by the addition of primers TSK8E and TAG to final concentration of 1µM. The reactions were cycled using a Perkin Elmer 4800^{TM} thermal cycler, each cycle consisted of denaturation at 96°C for 60 seconds, primer annealing at 67°C for 60 seconds and extension at 70°C for 180 seconds. The products were resolved by electrophoresis through a 1.5% Agarose gel (FMC), the was gel Southern blotted and hybridised to a radioactively labelled probe containing the sequence of the SINE immediately flanking the Xp:Yp telomere (generated by the PCR amplification of clone TSK8 (Royle, *et al* 1992 [redacted]) using primers TSK8A and TSK8B). Hybridisation was carried out using the 'blotto' hybridisation mix and the filter washed to a stringency of 0.5×SSC/0.1%SDS.

B. The same Southern blot as used in **A** hybridised to the telomere repeat probe which consists of a tandem array of TTAGGG repeats. The same products can be observed as observed in **A**. The strength of the hybridisation signal with the telomere repeat probe is reduced for the smaller products since these will contain less TTAGGG repeats. Dark smears can be observed from the hybridisation with the telomere repeat probe, probably due to hybridisation to single-stranded DNA generated from all the other telomeres. Hybridisation with the telomere repeat probe was carried using the phosphate/SDS hybridisation solution and the filter was washed to a stringency of 0.5×SSC/0.1%SDS.

C. An attempt to improve the resolution of the TVR-PCR products. TVR-PCR was carried using TelD and the flanking primer TSK8E on two unrelated individuals. The resolution of the TVR-PCR products was improved in the 40cm long agarose gel. Individual P53 was used in the experiments shown in A, B and C. Comparison of A and C shows the improved resolution of the TVR profile in DNA p53, the wide bands have been resolved into multiple bands which differ in size by approximately 6 bp.

The 7µl reactions contained 1µM of each primer. The reactions were cycled for 17 cycles with the same parameters as stated above. The TVR-PCR products were resolved by electrophoresis through a 40cm long gel containing 1% NuSieve agarose (FMC) and 1% agarose (FMC), the gel was Southern blot hybridised to a probe containing the SINE sequence as above.



Chapter 4. Figure 2

resolution was still not good enough to observe in detail every repeat unit, so that an unambiguous code could be generated. To further improve the resolution of these products, the analysis was carried out using denaturing polyacrylamide gel electrophoresis. The products were detected by PCR amplification with a radioactively 5′ end labelled flanking primer (TSK8E) followed by resolution of single stranded products on a denaturing polyacrylamide gel. Figure 4.3. shows one such denaturing polyacrylamide gel. In this experiment another telomere primer TelC was also used, it was designed to anneal to TTAGGG repeats (Table 4.1.). Using denaturing polyacrylamide gel electrophoresis the resolution was much improved. The four individuals analysed all showed different patterns in the distribution of the TTAGGG and TGAGGG repeats, again demonstrating the presence of a great deal of variation in the telomere repeat array (Fig. 4.3.). Many positions were observed where bands were present at the same position in the same individual in both the TelC and TelD reactions. At this stage in the development, the technique assayed the distribution of telomere repeats present in both Xp:Yp telomeres, so a diploid code was generated (Fig. 4.3.).

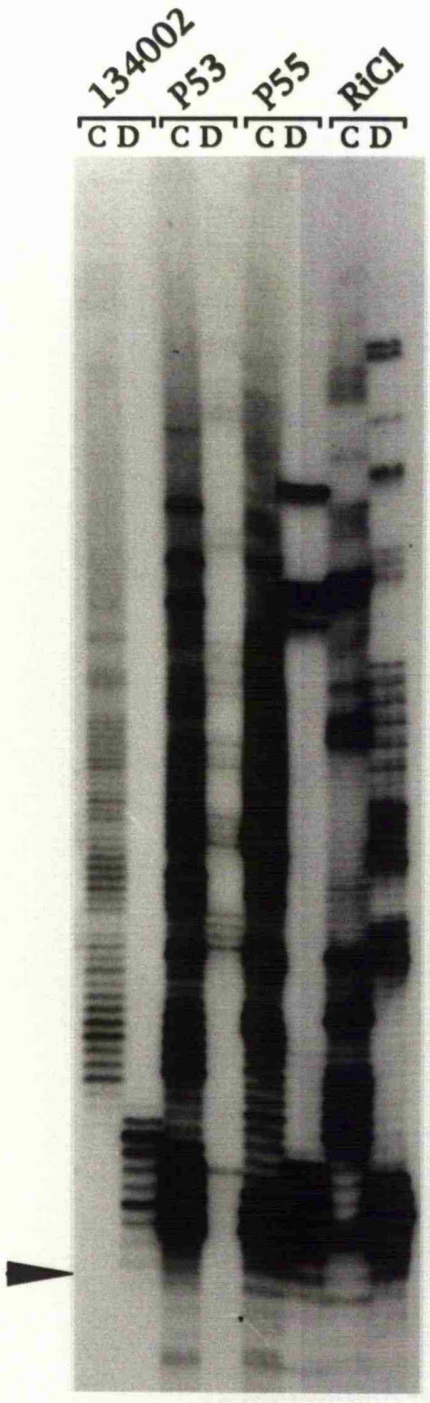
5. Use of flanking polymorphisms to generate haploid telomere repeat codes

The flanking primer TSK8E used in the above TVR-PCR experiments anneals approximately 410bp from the start of the telomere array. This caused problems in resolution of products that differ in size by 6bp. In addition, it was clear from previous experiments that the diploid TVR maps produced using TSK8E would be limited in their utility. This was because it was not possible to analyse a single telomere or read the code very far into the telomere, partly because occasional repeat unit length variants put the two allelic codes out of register in a diploid pattern. Potentially these problems could be over come, by utilising the sequence polymorphism in the DNA flanking the telomere (Chapter 3). Flanking primers were designed that were closer to the telomere repeat array, these primers TS-30A and TS-30T have a 3′ mismatch corresponding to alleles at the -30 polymorphic position (see chapter 3). Under optimal conditions these primers, in conjunction with the telomere repeat primers, should specifically amplify just one telomere from an individual heterozygous for the -30 polymorphic position. This should result in haploid telomere maps with products small enough to be resolved by denaturing polyacrylamide gel electrophoresis. Although a large amount sequence polymorphism was identified in the DNA flanking the Xp:Yp telomere (chapter 3), the strong linkage disequilibrium meant that the maximum heterozygosity of the flanking polymorphisms was roughly equal to that of a single diallelic polymorphism. Consequently the utility of these polymorphisms for allele specific amplification was reduced.

Figure 4.3. TVR-PCR with two telomere repeat primers, and the products resolved by denaturing polyacrylamide gel electrophoresis.

TVR-PCR was carried out with two telomere repeat primers TelC (C) and TelD (D) designed to anneal to TTAGGG and TGAGGG repeat respectively and the 5′ radioactively labelled flanking primer TSK8E. Four DNAs from unrelated individuals are shown, the pattern of each DNA is composed of two lanes the left hand lanes, C, contains products generated with TelC (TTAGGG) and the right hand lanes, D, contains products generated with TelD (TGAGGG). The individuals RiCl and P53 are the same as those shown in Figure 4.2c. The resolution of the TVR-PCR products is improved and shows a laddering with a 6 bp periodicity. Many bands appear in the same positions in both lanes (C and D) of a DNA because the distribution of telomere and variant repeats is assayed from both Xp:Yp telomeres simultaneously. The arrow on the left denotes the start of the telomere repeat array.

The 7μl PCRs contained 1μM of the telomere repeat primers and 0.1μM of 5′ radioactively end-labelled TSK8E, the standard PCR buffer (see chapter 2) and 0.1 U/μl of Taq polymerase. The reactions were cycled 17 times using a Geneamp 9600^{TM} thermal cycler (Perkin Elmer), each cycle consisted of denaturation at 96°C for 40 seconds, primer annealing at 68°C for 60 seconds and extension at 70°C for 50 seconds. To each PCR 5μl of 100% formamide solution containing xylene cyanol and bromophenol blue was added. 6.5μl of the PCR reaction was electrophoresed through a 1×TBE 5% denaturing polyacrylamide gel. The products were detected by autoradiography.



Chapter 4. Figure 3

6. Improving the telomere repeat primer design

The major technical difficulty in the design of the TVR-PCR system was the design of the primers used to detect the different types of telomere repeats. This is due to the short length of the telomere repeat (6bp). The telomere repeat primers must be long enough such that the optimal annealing temperature is similar to that of the flanking primer with which they are used. This is important since the locus specificity is based entirely on the flanking primer. The TelD and TelC telomere repeat primers were composed of nearly four identical repeats and therefore they anneal most efficiently to groups of four or more identical complementary repeats. As a consequence, repeat units in groups of three or less were not detected efficiently. To overcome these problems sets of new telomere repeat primers that included redundant positions were designed and tested.

The telomere repeat primers TelD and TelC were originally designed by Dr. Nicola Royle for use in a telomere anchored PCR strategy to isolate telomere junction fragments (Royle *et al.*, 1992) where minimal discrimination between TTAGGG and variant repeats was required. These primers were not designed to discriminate between TTAGGG and variant repeats at the 3′ most base and they contained a 5′ tail for cloning purposes. Having the mismatch at the 3′ end should result in greater discriminatory power and would allow the design of other primers to assay other repeat unit variants which differed at the second base of the repeat unit (TTAGGG, TGAGGG, TCAGGG and TAAGGG). Primers TelH and TelG were designed to discriminate between TTAGGG and TGAGGG repeats respectively. These primers discriminated between the repeat unit variants at the 3′ most base, had no 5′ tail and each consisted of four identical repeats (Table 4.1.).

Experiments were carried out to assess the performance of the new telomere repeat primers TelG, TelH and the new flanking primers TS-30A/T. Amplification with these primers in individuals heterozygous for the -30 polymorphic position, resulted in haploid telomere maps that were easily resolved on a 6% denaturing polyacrylamide gel (Fig. 4.4.). Although these telomere maps were improved it was still difficult to interpret them as a digital code because there were large gaps in the 6bp repeat ladders. The gaps probably occurred because of additional repeat unit variants that could not be detected by these primers. There were also many repeat positions where it was difficult to assign unambiguously the type of repeat that was present at that position, for example, the distal repeats of a block faded out making it difficult to define the end of the block of repeats (Fig. 4.4.). It was thought that the inability of these telomere repeat primers to detect repeats that occurred in groups of three or fewer contributed to this problem. However,

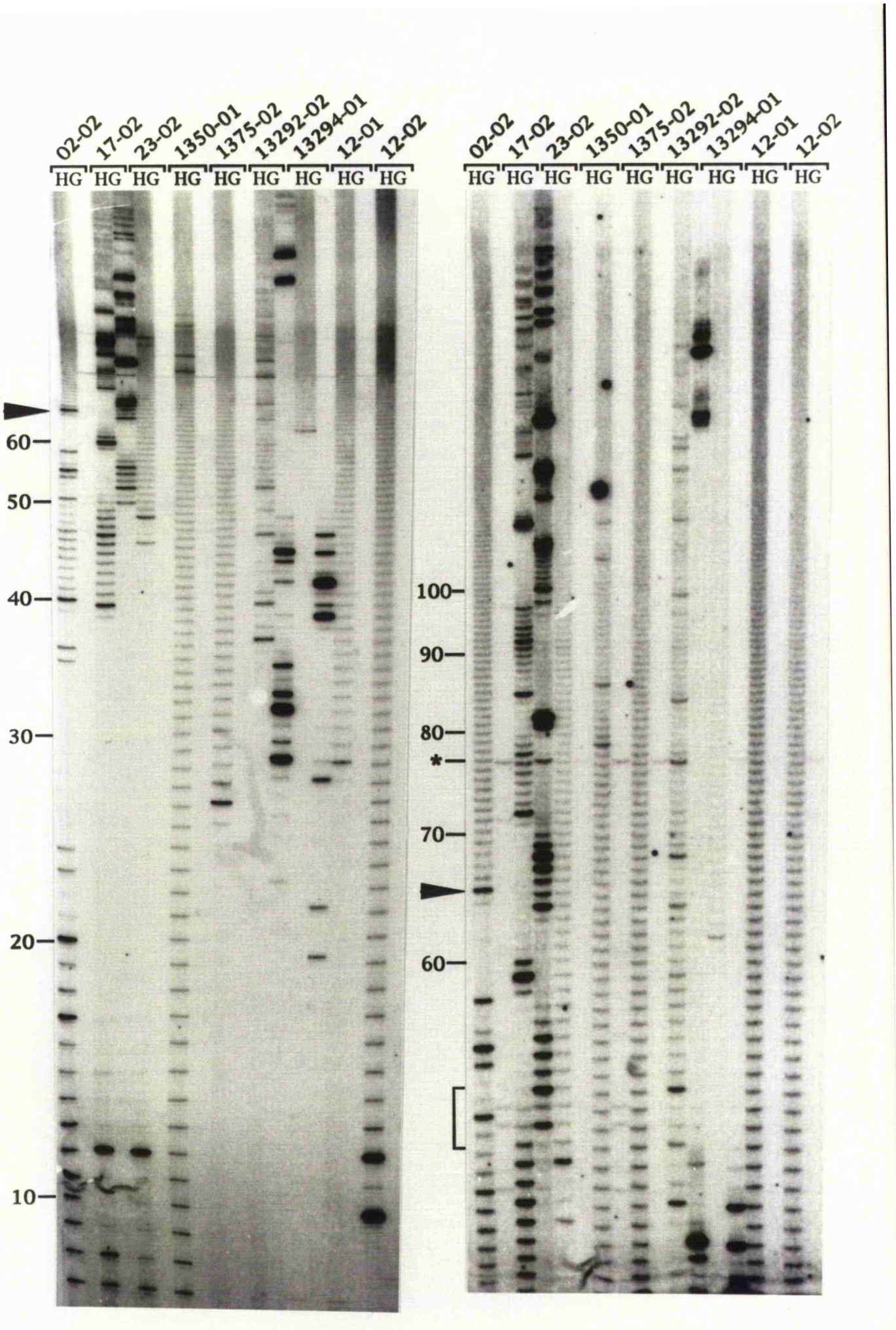
Fig. 4.4. Allele specific TVR-PCR on nine unrelated Caucasian individuals heterozygous at the -30 polymorphic position.

TVR-PCR on DNAs heterozygous at the -30 polymorphic position using the allele specific primer TS-30A and the telomere repeat primers TelH (H) and TelG (G) designed to anneal to TTAGGG and TGAGGG repeats respectively.

Two denaturing polyacrylamide gels of 6% and 5% are shown on the left and right respectively. The arrow to the left of the gel indicates the same sized product on both gels and the scale on the left side indicates the number of repeats from the start of the Xp:Yp telomere repeat array. Each DNA has two lanes H and G which show the distribution of TTAGGG and TGAGGG repeats respectively. Telomeres which contain large gaps are shown for example, 2302 which has a gap of 34 repeats between repeats 12 and 46. These gaps presumably contain additional telomere variant repeats to which the telomere repeat primers used can not anneal. The distal most repeats in a block of repeats fade out making it difficult to define the end of the block of repeats, for example, the repeats in individual 1702 between the bracket on the left of the 5% gel show the junction between the end of a block of TTAGGG repeats and the beginning of a block of TGAGGG repeats but it is not clear where these blocks of different repeats start and finish.

The asterisk marks the presence of a band which is present in all the TVR-PCR reactions with the TGAGGG repeat primer (TelG). These are thought to be amplification products from another locus as these products can be removed by the digestion of the input DNA with restriction enzymes which do not cut between the primer TS-30A and the Xp:Yp telomere.

The 10µl PCRs contained 1µM of the telomere repeat primers and 0.1µM of the radioactively 5´ end labelled allele specific flanking primer TS-30A, the standard PCR buffer (see chapter 2) and 0.1 U/µl of Taq polymerase. The reactions were cycled 19 times using a Geneamp 9600^{TM} thermal cycler (Perkin Elmer), each cycle consisted of denaturation at 96°C for 60 seconds, primer annealing at 68°C for 60 seconds and extension at 72°C for 150 seconds. To each PCR 5µl of 100% formamide solution containing xylene cyanol and bromophenol blue was added. 6.5µl of the PCR reaction was electrophoresed through a 6% denaturing polyacrylamide buffer gradient (0.5-2.5×TBE) gel and through a 5% denaturing polyacrylamide in 1×TBE gel. The products were detected by autoradiography.



Chapter 4. Figure 4

these primers demonstrated again that the distribution of TTAGGG and TGAGGG repeats was variable. Remarkably, specific amplification was achieved using the flanking primers TS-30A/T which are situated in the SINE. This may be because this family of SINEs have not been identified near telomeres and this copy of the SINE may be sufficiently diverged from other members of the family that the TS-30A/T primers are locus specific.

6.1. The use of inosine in telomere repeat primers. Additional telomere repeat primers were designed in an attempt to improve the detection of TTAGGG and TGAGGG repeats that occur singly within the telomere repeat array. Redundancy was incorporated in the three 5′ most repeats of the telomere primer. The base inosine was used since this had been reported to provide a 'inert' base which neither stabilised nor destabilised a duplex (Martin and Castro, 1985) and it has been used successfully in the design of degenerate PCR primers (Knoth *et al.*, 1988; Patil and Dekker, 1990; Wilkie and Simon, 1991). The use of inosine, instead of a mixture of bases, would reduce the complexity of the primers to a minimum. Two primers that annealed to the TGAGGG repeat type were synthesised; TelF had two redundant positions in the second repeat unit and inosine in the third and fourth repeat units (TelF: 5′-CCCIIA CCCIIA CCC(CT)(AC)A CCCTC-3′)(Table 4.1.). The redundant positions incorporated into the second repeat unit of TelF were designed to anneal to TTAGGG, TGAGGG and TTGGGG because it was thought that one of the most common variant repeat units would be TTGGGG (Allshire *et al.*, 1989). The second primer TelL was similar to TelF but contained inosine in the second repeat unit as well as the third and fourth repeat units (TelL: 5′-CCCIIA CCCIIA CCCIIA CCCTC-3′)(Table 4.1.). Both TelF and TelL worked as telomere repeat primers and they amplified from some repeats that the TelG primer had not, suggesting that they detected single TGAGGG repeats within the telomere repeat array. However, they often failed to amplify from repeats that the TelG primer had amplified (Fig. 4.5a.) and the TelL primer appeared to operate with less efficiency suggesting that the presence of inosine affected the thermal stability of the primer (Fig. 4.5a.).

6.2. Additional repeat unit variants. In order to investigate the nature of the gaps in the telomere maps, sequence analysis of telomeres in which the TVR codes contained large gaps was carried out. This revealed the presence of TCAGGG repeats. The presence of these variant repeats could have contributed to the failure of the TelF primer to amplify as expected as this primer was not designed with these repeats in mind. A telomere repeat primer, TelK, was designed to detect TCAGGG repeats (Table 4.1.). TelK consisted of a sequence complementary to four TCAGGG repeats and it successfully filled in many of the gaps in the telomere

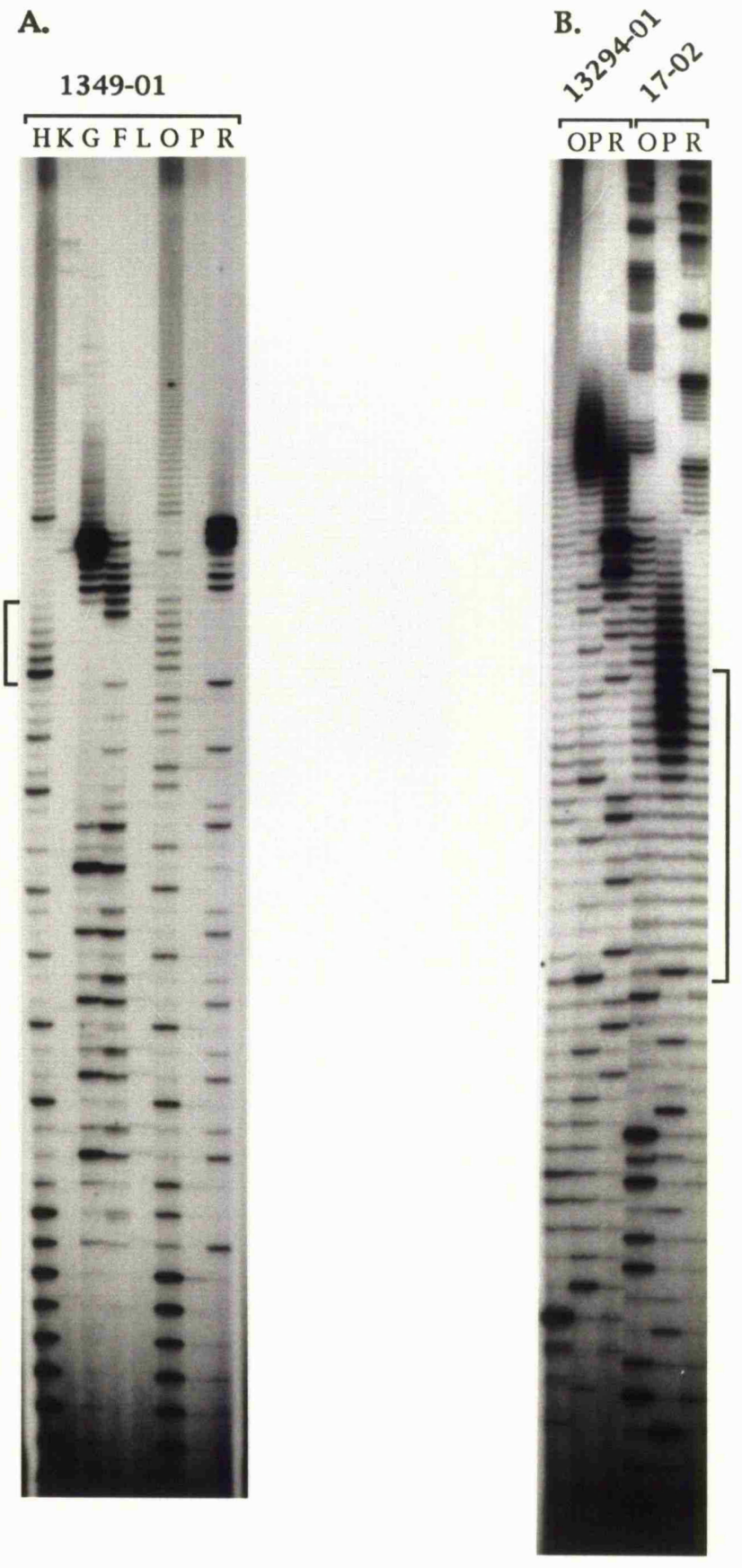
Fig. 4.5. Comparison of TVR-PCR with different telomere repeat primers.

A. TVR-PCR using primers TelH (H) (TTAGGG), TelK (K) (TCAGGG), TelG (G) (TGAGGG), TelF (F) (TGAGGG), TelL (L) (TGAGGG), TelO (O) (TTAGGG), TelP (P) (TCAGGG) and TelR (R) (TGAGGG). Descriptions of these primers are the text of chapter 4 and the sequences are in Table 4.1. DNA from the Caucasian 1349-01 was used, this individual is heterozygous at the -30 polymorphic position flanking the Xp:Yp telomere. Comparison of the patterns generated by the primers TelR with TelG, TelF and TelL, all of which are designed to detect TGAGGG repeats, revealed that TelR produced the best patterns where single repeats were identified. The beginning and end of blocks of repeats are defined by the TelO and TelR primers, for example, compare the block of repeats bracketed on the left in the TelH lane with the same block in the TelO lane. The telomere shown contains no TCAGGG repeats, so the lanes containing products from the amplification by the TelK and TelP primers contained no products.

The 10μl PCRs contained 1μM of the telomere repeat primers and 0.1μM of the radioactively 5′ end labelled allele specific flanking primer TS-30A. The reaction conditions and cycling parameters were the same as described for Fig. 4.4.

B. TVR-PCR with telomere repeat primers TelO, TelP and TelR in two heterozygous individuals with the allele specific flanking primer TS-30A. Both the telomeres shown contained TCAGGG repeats and demonstrated the inability of the TCAGGG repeat primer, TelP, to discriminate between the different repeat types. This also demonstrated the inability of the TTAGGG and TGAGGG repeat primers TelO and TelR to discriminate against amplification from TCAGGG repeats. For example, the region in the telomere 17-02 within the bracket (right) appears to contain TCAGGG repeats but the start and finish of this block is not defined with any of these primers (compare with the 17-02 telomere map generated by amplification with TelY in Fig. 4.6b.). Primers TelO and TelR also generate products in this region and the beginning and end of the block of TCAGGG repeats is not defined by the TCAGGG repeat primer TelP. The reaction conditions and cycling parameters were identical to those described above.

These figures demonstrate that the telomere repeat primers TelO and TelR can discriminate between TTAGGG and TGAGGG repeats but the presence of TCAGGG repeats can not be correctly defined by TelO, TelP or TelR.



Chapter 4. Figure 5

maps but it still did not perform well enough to produce an easily interpretable code.

6.3. Further telomere repeat primer design. A new set of telomere repeat primers were synthesised TelO, TelP and TelR designed to detect TTAGGG, TCAGGG and TGAGGG repeats respectively (Table 4.1.). These were all identical expect for the 3′ most base which was designed to discriminate between the repeat types. These primers contained a redundant position (N) in the second repeat unit and inosine in the third and fourth repeat units so keeping the complexity of the primer pool to a minimum (TelP: 5′-CCCTIA CCCTIA CCCTNA CCCTG). These primers performed much better than previous telomere repeat primers (Fig. 4.5a.). The first and last repeats of an array of identical repeats were more precisely defined and single repeat units were detected so the information content of the telomere repeat maps was increased (Fig. 4.5a.). However, these repeat primers only performed well on telomeres that did not contain any TCAGGG repeats (such as 134901, Figure 4.5a.). In more complex telomeres containing TCAGGG repeats, overlap in the pattern produced by each primer could be seen, making the reliable deduction of a telomere code impossible (Fig. 4.5b.). In telomeres containing TCAGGG repeats, the TelP primer would identify the presence of TCAGGG repeats but additional TCAGGG repeats were synthesised at the end of blocks of TCAGGG repeats (Fig. 4.5b.) and this problem increased with increasing PCR cycle number. It was also noted that the telomere repeat primers TelO and TelR that amplified from TTAGGG and TGAGGG respectively, also primed synthesis at the TCAGGG repeat positions (Fig. 4.5b.). These effects probably arose because inosine did not base pair with A, C, G or T equally. Instead, in these assays as in others, inosine mimics guanine so pairing with C with a greater affinity than with A, G or T (Martin and Castro, 1985). It has been reported that Taq polymerase recognises inosine as guanine and inserts a cytosine on the complementary strand (Wilkie and Simon, 1991). This would account for the additional bands that were synthesised in the TVR-PCRs containing TelP as this primer was acting a template for the synthesis of more TCAGGG repeats in each round of PCR. The ability of inosine to base pair more efficiently with cytosine would also account for the ability of TelO and TelR (TTAGGG and TGAGGG) to prime synthesis from TCAGGG repeats.

It was clear that although the incorporation of degeneracy using inosine produced an improvement in the ability to code telomere repeat arrays devoid of TCAGGG repeats, the use of inosine would not be practical for all telomere repeat arrays and an alternative primer design had to be found.

6.4. The use of thymidine in the telomere repeat primers. Instead of trying to find a base that would pair with a similar thermal stability with all four bases, it was decided to use a base that would not base pair with any of the bases likely to be present in the second position of the telomere and variant repeats. Although such a base would lower the thermal stability of any duplex formed, it was hoped that it would do so equally no matter which base was present. Sequence analysis of cloned telomeres had revealed that few TAAGGG repeats are found in human telomeres (N. J. Royle pers. comm.) A primer designed to detect such repeats was synthesised (TelT), this primer did not detect any TAAGGG repeats in the Xp:Yp telomere repeat arrays of ten individuals. As a result thymidine was used in place of inosine in the third and fourth repeat units of the telomere repeat primers.

A new set of telomere repeat primers were synthesised, all identical except at the 3′ end. TelW, TelX and TelY were designed to detect TTAGGG, TGAGGG and TCAGGG repeats respectively (TelW: 5′-CCCTTA CCCTTA CCCTNA CCCTA-3′)(Table 4.1.). These telomere repeat primers performed well, the beginning and ends of arrays of identical repeats were well defined, single repeat units were easily identified and the discrimination of all the primers was good (Fig. 4.6a.). The ends of arrays of TCAGGG repeats showed a small amount of additional laddering. Due to the deliberate mismatch caused by the presence of a T in the third and four repeats, these three primers have a lower thermal stability compared to the other telomere repeat primers used resulting in a reduction of the sensitivity (Fig. 4.6a.). It was thought that the use of the 5′ TAG sequence with the use of the TAG primer would provide the additional sensitivity required, since amplification would be uncoupled from detection (Jeffreys *et al.*, 1991). The primers were re-synthesised with the 20 base 5′ TAG sequence, TAG-TelW, TAG-TelX and TAG-TelY (Table 4.1.). The uniformity of the pattern was much improved, particularly for single repeat units and the small amount of laddering at the end of arrays of TCAGGG repeat was eliminated (Fig. 4.6b.) possibly due to an 'anchoring' effect of the 5′ TAG sequence. However, the use of the TAG primer itself in the PCR reaction made no difference to the sensitivity of the system (data not shown). This suggested that the presence of the 20 base 5′ tail (TAG) and not the TAG primer was responsible for the increased sensitivity, possibly due to an increase in the thermal stability of these primers so more efficient amplification could be achieved. This may appear contradictory to the initial experiments where the presence of the TAG sequence and the use of the TAG primer decreased the efficiency of the reaction, this could be because these experiments did not test the utility of the TAG primer itself. It appears that in this system the presence of the 5′ TAG sequence increases the efficiency of the reaction but including the TAG primer makes no difference to the reaction. This effect has been noted in an

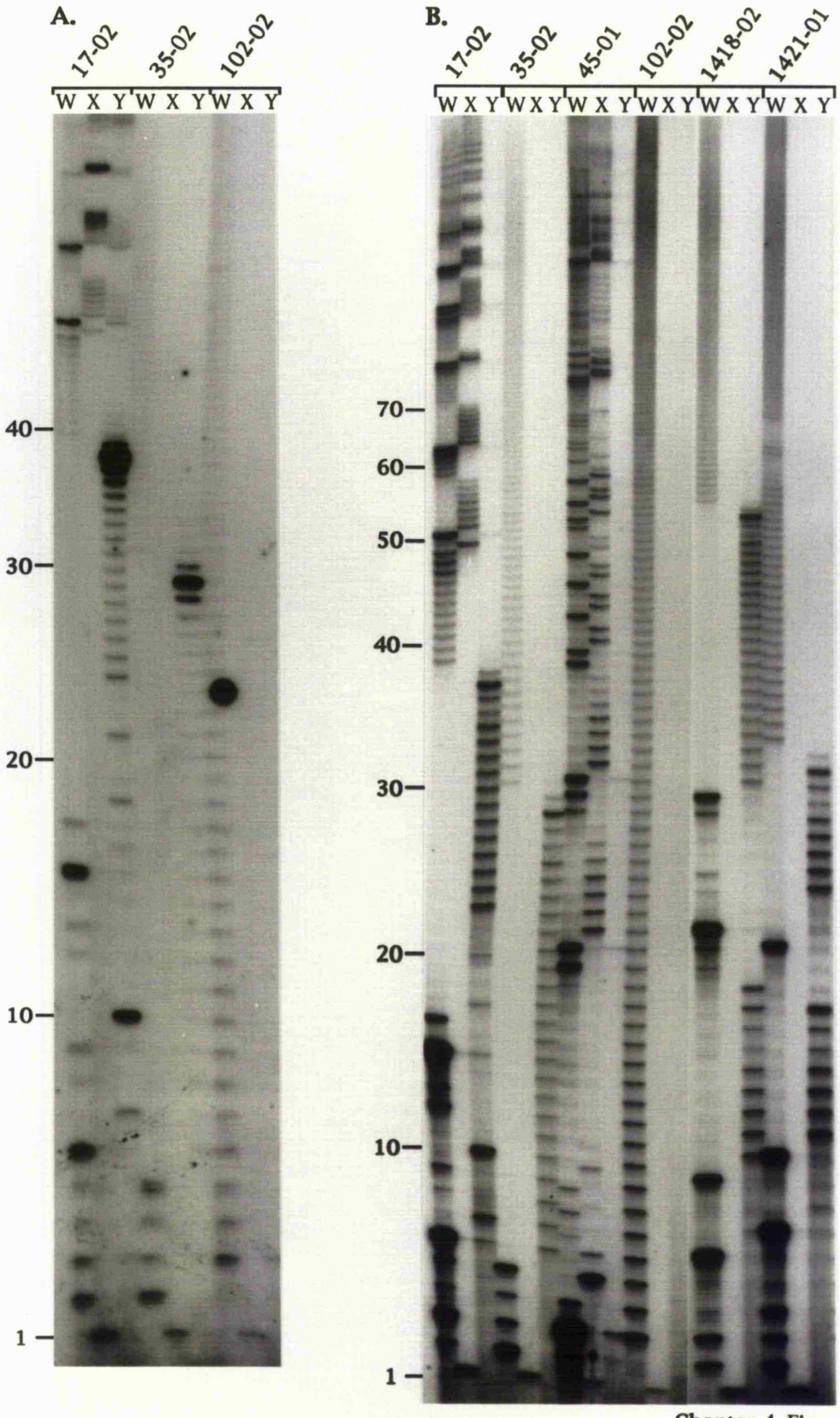
Fig. 4.6. TVR-PCR with telomere repeat primers TelW, TelX and TelY with and without the 5′ TAG sequence.

A. TVR-PCR with the telomere repeat primers TelW, TelX and TelY designed to detect TTAGGG, TGAGGG and TCAGGG repeats without the 5′ TAG sequence. DNA from three unrelated individuals heterozygous at the -30 polymorphic position were used. Although the definition of the telomere repeat ladder is good the signal is not strong.

The 10µl PCRs contained 0.4µM of the telomere repeat primers and 0.1µM of the radioactively 5′ end labelled allele specific flanking primer TS-30A, the standard PCR buffer (see chapter 2) and 0.1 U/µl of Taq polymerase. The reactions were cycled 19 times using a Geneamp 9600^{TM} thermal cycler (Perkin Elmer), each cycle consisted of denaturation at 96°C for 20 seconds, primer annealing at 67°C for 40 seconds and extension at 70°C for 120 seconds. To each PCR 5µl of 100% formamide solution containing xylene cyanol and bromophenol blue was added. 6.5µl of the PCR reaction was electrophoresed through a 6% denaturing polyacrylamide gel with a 0.5-2.5×TBE buffer gradient and the products detected by autoradiography.

B. TVR-PCR with telomere repeat primers with the 5′ TAG sequence TAG-TelW, TAG-TelX and TAG-TelY. DNA from six unrelated individuals heterozygous at the -30 polymorphic position were used. Individuals 17-02, 35-02 and 102-02 are same as those used in **A**. The definition of bands is even better and the strength of the signal is much improved.

The reactions were carried out using identical conditions to those described above.



Chapter 4. Figure 6

MVR-PCR system at the Y chromosome specific minisatellite MSY1 (M. Jobling pers. comm.).

The telomere repeat primer design has been optimised and the TAG-TelW, TAG-TelX and TAG-TelY primers in conjunction with allele specific flanking primers can be used to assay the distribution of TTAGGG, TGAGGG and TCAGGG repeats in single Xp:Yp telomeres. The position of the repeat types can be reproducibly interpreted as a digital telomere repeat code, see figure 4.7.

7. Confirmation of telomere repeat codes.

To confirm that the telomere codes generated from the TVR-PCR technique reflected the repeats actually present in the telomere repeat array, sequence analysis was carried out on bands cut from a dried TVR-PCR polyacrylamide gel. The sequence of the telomere repeats present matched that derived by TVR-PCR with the exception of additional variant repeats, usually those that caused a change in repeat length. The presence of such repeats resulted in no amplification from the repeat proximal to the length variant repeat; the non-amplifying repeat was assigned N (Fig. 4.8.). Single variant repeats such as TTTAGGG were detected by primer TAG-TelW and typed as T-type repeats. For example, the ninth repeat from the start of the telomere in individual 140801 was a TTTAGGG repeat it was detected by the TAG-TelW primer designated T in the telomere repeat code (Fig. 4.8.).

8. Confirmation that the telomere repeat codes originate from a single telomere.

Given that the telomere primers can prime synthesis from all 92 telomeres present in a diploid cell, and that the flanking primers TS-30A and TS-30T are situated in a SINE which is estimated to be present in 20,000 copies in the human genome (La Mantia *et al.*, 1989), it was necessary to demonstrate that telomere maps were generated from a single telomere. Observations during the course of the development of this technique suggest that amplification from more than one telomere was unlikely. The telomere pattern observed when using primer TSK8E, which was not located in the SINE, was similar to that observed when the TS-30A/T primers were used. The use of allele specific primers produced a telomere repeat pattern concordant with that expected from a single telomere from individuals heterozygous for the -30 polymorphism in the flanking DNA. Individuals homozygous for the -30 polymorphism created diploid telomere maps with one primer while the other primer failed to amplify. However, additional bands that did not fit into a haploid telomere map were observed, these bands

Fig. 4.7. Showing the TVR-PCR technique.

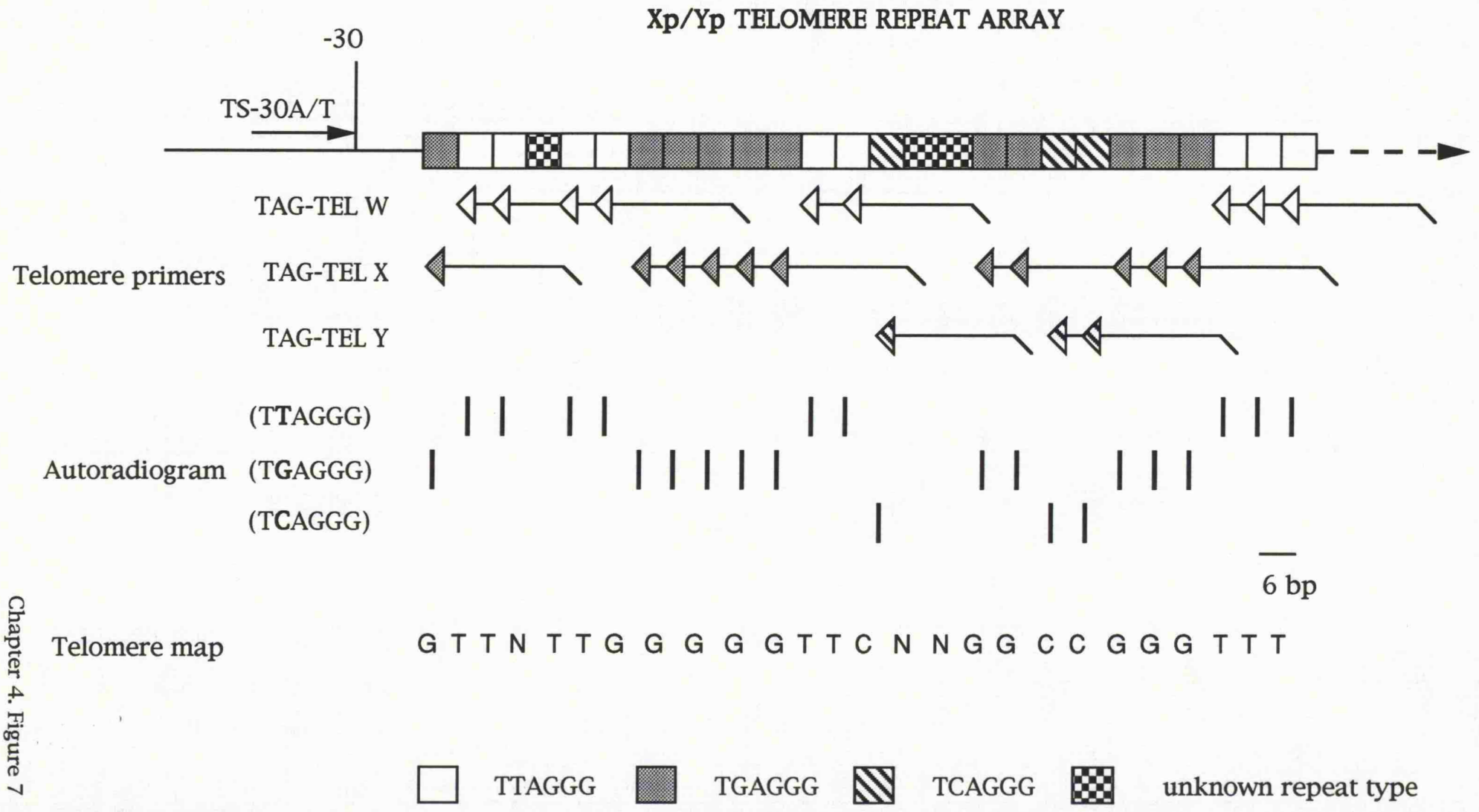


Fig. 4.8. Telomere repeat codes compared to the sequence of the telomere.

Individual 140801. Flanking primer TS-30A.

	flanking DNA	Telomere repeat array	TAG-TelY
Sequence	ggtccactttca	gagggttagggttagggttagggttagggtttagggtcagggttaggggttagggtcagggtcagggtcagggtaagggtaagggtcc..	
Telomere repeat code		G T T T N T C N T C C C	

Individual 134601. Flanking primer TS-30T.

	flanking DNA	Telomere repeat array	TAG-TelX
Sequence	ggtccactttca	gagggtgaaggtgagggtgagggtgagggtgagggtgagggtgagggtaagggtaagggtcc..	
Telomere repeat code		N N G G G G G G	

The sequence of the proximal ends of single Xp:Yp telomere repeat arrays from two individuals of the CEPH panel of DNAs compared to the codes obtained by TVR-PCR. T=TTAGGG, G=TGAGGG, C=TCAGGG and N=non-amplifiable repeat positions. The underlined sequences show the sequence of the telomere repeat primers. Some positions are assigned N by TVR-PCR, but the sequence shows that some of these repeats are a type of repeat unit which the telomere repeat primers are designed detect. These repeats do not amplify because of an additional variant repeat immediately distal, such as TGAAGG at the second repeat of individual 134601 which results in the first position GAGGG being assigned N.

could be derived from additional loci (as observed in Fig. 4.4.). The double digestion of the input genomic DNA with two frequently cutting restriction enzymes (*AluI* and *DdeI*) that did not cut between the Xp:Yp telomere and primers TS-30A/T, removed these additional bands. Finally and most conclusively, the telomere maps were observed to segregate through pedigrees in a Mendelian fashion. The segregation in part of one of the seven CEPH pedigrees analysed is shown in figure 4.9.

9. Steps towards automation

The TVR-PCR technique involves the use of ^{32}P 5′ radioactively end labelled primers and polyacrylamide gels. The maximum capacity of most commercially available polyacrylamide gel apparatus is approximately 16 telomere maps per gel and most of the telomere maps produced require additional resolution of larger fragments on a lower percentage gel. A method that obviates the need for radioactivity and allows the resolution of a complete telomere map on a single gel would be welcome. An automatic DNA sequencing machine utilises a fluorescence detection method and if the correct gel concentration is used could resolve the fragments generated by TVR-PCR. These have been utilised for the analysis of short tandem repeat loci by multiplex PCR using different coloured fluors (Reed *et al.*, 1994). The system for the analysis of fluorescently labelled PCR products on an Applied Biosystems sequencing machine is called a 'gene scanner'. The TVR-PCR technique was adapted for use on a gene scanner. This involved the synthesis of three differently labelled flanking primers, the TS-30A primer was labelled with each of the fluorescent dyes HEX, 6-FAM and TET (Applied Biosystems). A labelled flanking primer was used with one of the telomere repeat primers, the PCRs were carried out separately and then combined and loaded into a single lane on the gene scanner gel. An electropherogram was produced and a ladder of bands of three different colours was observed, a different colour for each telomere repeat type assayed. This allowed one telomere to be analysed per lane of the gel, on the ABI 373 machine 36 telomeres can be analysed per gel. However, the technique was less sensitive than the manual method with a ^{32}P radioactively end-labelled flanking primers. Only after 25 cycles of PCR was the signal strong enough to be detected by the ABI gene scanner and converted to a telomere repeat code (Fig. 4.10.). This technique will be developed further to optimise the conditions and to increase the sensitivity. The data from telomere mapping presented in this thesis were produced using ^{32}P radioactively 5′ end labelled PCR primers.

Fig. 4.9. The Mendelian segregation of two Xp:Yp telomeres through a pedigree

Two telomeres can be observed segregating through this pedigree. The grandfather (11) is heterozygous at the -30 polymorphic position and the telomere associated with the A allele at the -30 position is observed (A´). The grandmother (12) is homozygous for the A allele at the -30 polymorphism and two telomeres can be observed superimposed (A and A´´). The father (01) is also homozygous for the A allele at the -30 polymorphism and two superimposed telomeres can be observed, A´ from the grandfather and the A from the mother. Two children are shown both heterozygous at the -30 polymorphism. Single telomeres can be observed, the son (03) has the A´ telomere originally observed in the grandfather and the daughter (08) has the A telomere originally observed in the grandmother.

TVR-PCR reactions were carried as described in figure 4.6.

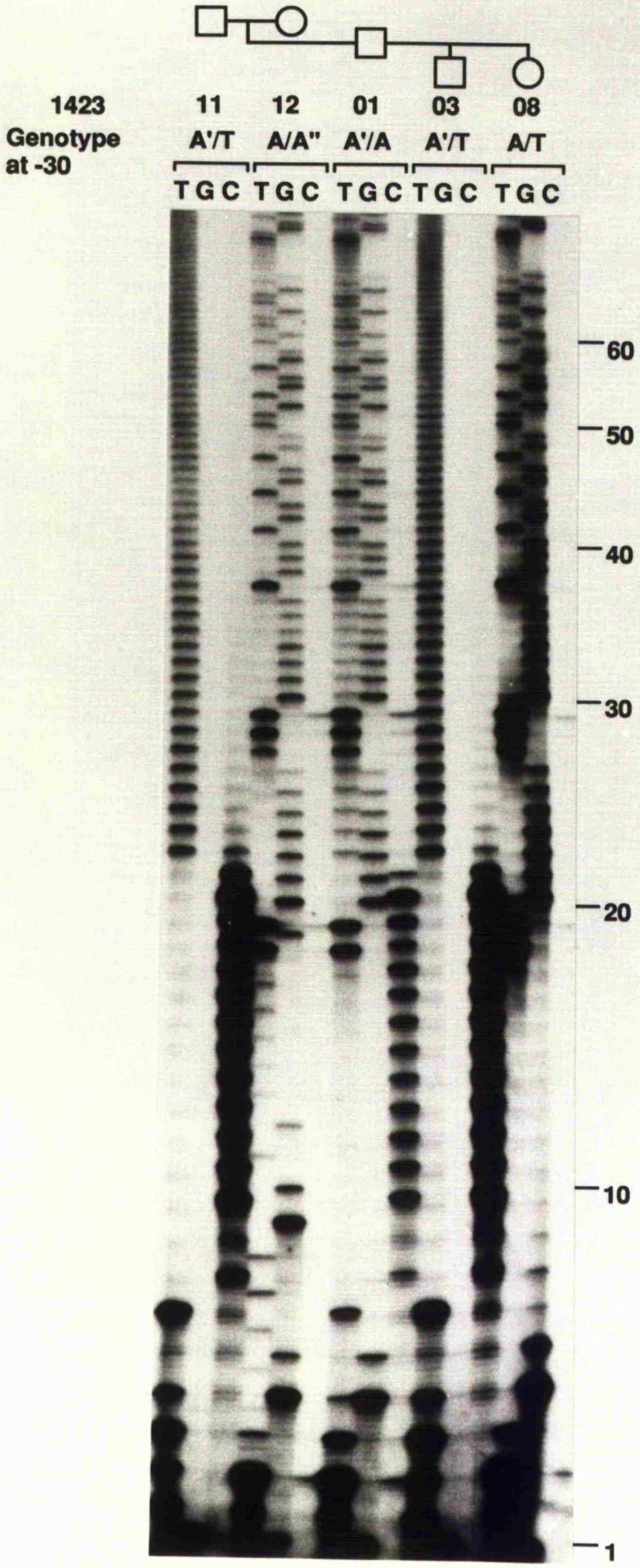
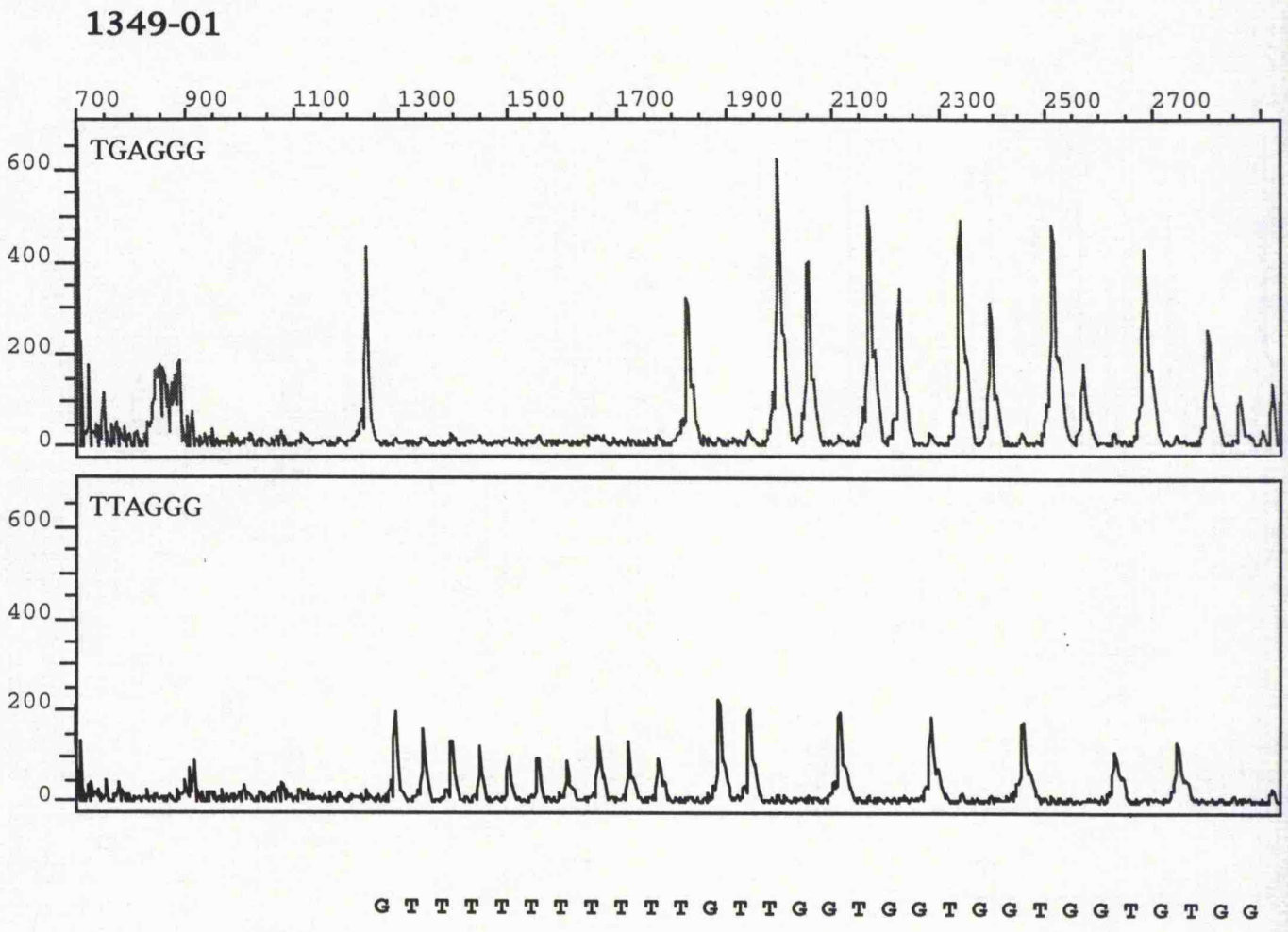

1423
Genotype
at -30
11
12
01
03
08
A'/T
A/A"
A'/A
A'/T
A/T
T G C T G C T G C T G C T G C
60
50
40
30
20
10
1

Fig. 4.10. TVR-PCR on an ABI GeneScaner.

Electropherograms of TVR-PCR products resolved on an Applied Biosystems (ABI) 373 automated sequencing machine running the gene GeneScan software. The electropherogram at the top represents the products generated with the TAG-TelX (TGAGGG) and the flanking primer TS-30A labelled with the ABI fluorescent dye 6-FAM. The electropherogram at the bottom represents products generated with the telomere repeat primer TAG-TelW (TTAGGG) and TS-30A labelled with the ABI dye HEX. These electropherograms would normally by superimposed and the signals from the different dyes distinguished by their different fluorescence (HEX= green; 6-FAM= blue). A code corresponding to the distribution of TTAGGG and TGAGGG repeats can be read from these electropherograms and is shown below the diagrams. The individual 1349-01 is the same as that shown in Fig. 4.5a.

The TVR-PCR reactions were carried out as described in Fig. 4.6. except that they were cycled 25 times, each reaction contained a differently labelled flanking primer and a different telomere repeat primer. The reactions were pooled and 2μl loaded onto a 5% denaturing polyacrylamide ABI 373 gel (see chapter 2).



Chapter 4. Figure 10

Discussion

In this chapter I have described the development of a technique to assay the distribution of TTAGGG, TGAGGG and TCAGGG repeats in the proximal 720bp of the Xp:Yp pseudoautosomal telomere repeat array. The TVR-PCR technique results in a ladder of products that can be interpreted as a code. Each code can be incorporated into a databank. At present this databank is simply a document containing all the codes (see chapter 5), but, as the number of codes increases it is likely that a more sophisticated databasing system will be required.

Forensic analysis. The distribution of the telomere and variant repeats at the proximal end of the Xp:Yp telomere is highly polymorphic. This raises the possibility that TVR-PCR at the Xp:Yp telomere could be utilised for individual identification, parentage analysis and in forensic analysis. TVR-PCR would be most useful if it was developed on a fluorescence based system such as the ABI gene scanner. This would remove the need for radioactivity and increase sample throughput. Further automation may be possible if direct computerised interpretation of the telomere repeat code could be achieved. However, the usefulness of TVR-PCR is reduced as a tool in forensic analysis, because it can only be carried out by allele specific amplification. This relies on the informativeness of the polymorphisms in the flanking DNA. The single base substitutional polymorphisms in the flanking DNA are in strong linkage disequilibrium (see chapter 3), therefore the maximum heterozygosity will depend on the frequency of the flanking haplotype in the population, but cannot exceed that of a single diallelic polymorphism (50%). However, this problem could be overcome if computer systems were developed for the analysis of diploid TVR-PCR electropherograms, such that they could be compared and similar codes identified, this would remove the reliance of the system upon the informativeness of the polymorphisms in the flanking DNA. TVR-PCR could be used to complement the powerful MVR-PCR systems at minisatellite loci. It is thought that several of these systems would have to be used to overcome problems that occur with the high underlying mutation rate of these loci. This could result in exclusions in parentage analysis that were due to mutation rather than a genuine difference. It is possible that TVR-PCR may be used in some situations to complement these systems and provide the additional power to overcome these problems. TVR-PCR may provide a useful system for the analysis of degraded DNA, since only a small region of DNA is analysed but contains a large amount of information.

Analysis of the dynamics of telomeres in somatic tissues. The TVR-PCR technique is based on the polymerase chain reaction which requires few input

molecules and therefore could be very sensitive, though the sensitivity of this technique has not yet been determined. It is hoped that TVR-PCR would be sensitive enough to assay the distribution of telomere repeats from small numbers of cells harvested from primary cell lines. These cell lines have a finite life span and undergo cellular senescence. Cellular senescence has been correlated with the reduction in length of the telomere array, such that when the cells reach crisis the telomere arrays are too short to protect the chromosome ends from degradation and fusion to other chromosomes (Counter *et al.*, 1992). Telomere repeat arrays have been observed that contain informative patterns of variant repeat interspersion. Such telomeres would provide a useful system to observe the dynamics of telomeres in senescing cell lines. Alterations in telomere length and chromosome fusions have been observed in tumour tissues, TVR-PCR on informative telomeres could provide a system to dissect such changes in telomere structure. Instability at short tandem repeat loci has been observed in a class of tumours termed RER+. TVR-PCR may provide a useful system for the detailed analysis of instability a six bp repeat arrays in such tumours.

This technique can be utilised to analyse variation at other human telomeres, work is in progress in this laboratory to develop a TVR-PCR system at the telomere of 7q (J. Coleman pers. comm.). TVR-PCR is in principal applicable to the telomeres of other organisms.

Conclusions

I have demonstrated that it is possible to assay the distribution of telomere and variant repeats in single Xp:Yp pseudoautosomal telomeres. TVR-PCR should provide a very useful system for the detailed analysis of the dynamics of telomere turnover both in somatic and in germinal tissues.

The analysis of variation and mutation in the Xp:Yp pseudoautosomal telomere repeat array in several human populations is presented in chapter 5.

Chapter 5

VARIATION IN THE Xp:Yp PSEUDOAUTOSOMAL TELOMERE REPEAT ARRAY

Summary

In this chapter I have described the analysis of sequence variation in the Xp:Yp telomere repeat array. The variation has been revealed using telomere variant repeat mapping by PCR (TVR-PCR), the development of which was described in chapter 4. The analysis revealed high levels of allelic variation in four human populations. A total of 229 single telomere repeat codes were derived from unrelated Caucasian, Japanese, West-African and Zimbabwean individuals. The vast majority of the codes were different, with the exception of 28 alleles which fell into three groups of identical alleles. In two of these groups (n=19 and 7) the alleles are virtually devoid of variant repeats, the third group contained two alleles from the same population which had identical telomere repeat codes. This variation results in heterozygosities estimated to be in excess of 99% in all the populations analysed, which is comparable to the most hypervariable minisatellites. The codes are grouped according to the haplotype in the flanking DNA (chapter 3). The analysis of these haplotypic groups suggests that the strong linkage disequilibrium observed in the DNA flanking the telomere (chapter 3), extends into the first few repeats of the telomere array. It appears that these telomeres have evolved along haploid lineages and that the predominant mechanism generating the high levels of variation must be intra-allelic. Comparison of related alleles showed that the majority of differences could be accounted for by small expansions or contractions in blocks of identical repeat units. However, some telomere repeat motifs were shared between the haplotypic groups, suggesting that occasional inter-allelic exchanges have contributed to the turnover of telomere repeats. In order to measure the mutation rate directly and infer the mutational mechanisms that generate the observed variability, a scan for mutations within pedigrees was carried out. One male mutation event was observed in a total of 160 male and female meiosis, this represents a mutation rate of 6.25×10^{-3} per gamete. This mutation consisted of a two repeat unit deletion in a block of 15 identical repeat units. The TVR-PCR analysis also revealed that there is a rare 3bp polymorphism in the start site of the telomere repeat array.

Introduction

In chapter four I described the development of a technique (TVR-PCR) to analyse the distribution of TTAGGG, TGAGGG and TCAGGG repeats in the Xp:Yp telomere repeat array. The analysis was carried out on single telomeres by allele specific amplification in individuals heterozygous for polymorphisms identified in the DNA flanking the telomere. During the course of the development of this technique it became clear that the distribution of the telomere and variant repeats was highly polymorphic. In order to quantify this variation I applied the TVR-PCR technique to human populations and created a data set of telomere repeat codes. This data set has allowed the extent of allelic variation to be quantified for the Xp:Yp telomere. Prior to this analysis, the variant repeat distribution had only been determined at all chromosome ends simultaneously (Allshire *et al.*, 1989). Telomere repeat codes showing similarities can be arranged into groups of related alleles which exhibit common ancestry. Comparison of alleles within and between groups will allow inferences to be made about the nature of the mechanisms that have influenced their evolution and generated the observed variation.

Mechanisms of tandem repeat turnover. Analysis of the distribution of different types of repeat unit has been carried out at minisatellite loci through the use of MVR-PCR (Jeffreys *et al.*, 1991). Comparison of groups of related minisatellite alleles revealed a polarity in variation and therefore mutation towards one end of the minisatellite array (Armour *et al.*, 1993; Neil and Jeffreys, 1993). The direct analysis of structural changes occurring during *de novo* mutations has been carried out by the MVR-PCR analysis of allele length mutants derived from pedigrees and directly from sperm DNA (Jeffreys *et al.*, 1994; Monckton *et al.*, 1994). Mutation events also show a strong polarity toward one end of the repeat array, both inter and intra-allelic mechanisms appear to be involved and the lack of exchange of flanking markers suggests the involvement of complex gene-conversion like events (Jeffreys *et al.*, 1994).

The best studied satellite DNA is α-satellite DNA found at the centromeres of all human and primate chromosomes (Manuelidis, 1978) and consist of a 171bp repeat unit arranged in long tandem arrays of several megabases in length (reviewed by Willard, 1990). The repeat units of α-satellite DNA appear to have a higher sequence similarity with repeat units from the same species than those from closely related species, suggesting that these sequences undergo concerted evolution (Durfy and Willard, 1990). Mechanisms to explain concerted evolution include repeat rounds of unequal exchange or conversion. The 171bp α-satellite monomer is arranged into chromosome specific higher order structure, probably

the result of exchanges between misaligned chromosomes (Willard and Waye, 1987; Greig *et al.*, 1993). This could also account for the large length polymorphisms observed in α-satellite DNA.

Two mechanisms have been proposed to account for the instability in simple repetitive DNA; unequal crossing over (Smith, 1973) and replication slippage (Streisinger *et al.*, 1973). Mutations that eliminate recombination in *E. coli* and *S. cerevisiae* do not affect instability (Levinson and Gutman, 1987; Henderson and Petes, 1992), mutations that affect DNA mismatch repair systems increase instability (Levinson and Gutman, 1987; Strand *et al.*, 1993; Ionov *et al.*, 1993). In addition, variability at human short tandem repeat loci including trinucleotide repeat loci implicated in several genetic disorders is proportional to perfect repeat copy number (Weber, 1990; Richards and Sutherland, 1992; Eichler *et al.*, 1994). The information described above suggests that the most likely mechanism involved in the instability of simple tandem repeats is replication slippage rather than recombination.

Variation in the start site of telomeres. Extraordinary length polymorphism has been detected at the ends of human chromosomes. Four alleles have been detected at the 16p telomere, these alleles place the α-globin genes 170, 245, 350 or 430kb from the telomere (Wilkie *et al.*, 1991; Harris and Thomas, 1992). These alleles are associated with different subterminal sequences. The most likely explanation for the existence of these alleles is that they arose by exchanges between subterminal sequences of non-homologous chromosomes. The alleles at 16p are in complete linkage disequilibrium with proximal markers, suggesting that there may be a suppression in recombination. This indicates that homologue recognition must occur interstitially or just proximal to these markers (Wilkie *et al.*, 1991). Chromosome length polymorphisms have been identified at the termini of other human chromosomes (Ijdo *et al.*, 1992; Harris and Thomas, 1992; Macina *et al.*, 1994). The Xp:Yp pseudoautosomal region is unique in the human genome and does not share the subterminal sequences present at the ends of many other human chromosomes. Since the observed polymorphism in the start site of telomeres is probably due to exchanges between subterminal repeats, one might expect that there would not be any polymorphism in the start site of the Xp:Yp telomere.

TVR-PCR will provide a method to analyse in detail the dynamics of a single telomere within human populations. This may result in information which will help to elucidate the types of mechanisms that result in telomere repeat turnover, whether inter-allelic, intra-allelic or exchanges between non-homologous chromosomes are involved. It will also provide a useful system for the study of

STR loci in light of the recent discovery of dynamic mutation in trinucleotide repeat arrays implicated in inherited disease (Richards and Sutherland, 1992) and instability at STRs in RER+ tumours (Aaltonen *et al.*, 1993). The study of variation with the type of technique which has been so successful at minisatellite loci (Jeffreys *et al.*, 1994) at a six bp repeat locus may complement research on shorter repeat unit loci. This could help bridge the gap in knowledge between tetranucleotide repeats and minisatellites. The analysis of variation in the telomere repeat array and in the flanking DNA (Chapter 3) will provide some information about possible chromosome length variation which had been demonstrated at other chromosome ends (Wilkie *et al.*, 1991; Macina *et al.*, 1994).

This work. In this chapter I have described the characterisation of allelic variation in the distribution of telomere and variant repeat units at the Xp:Yp telomere. This analysis was carried out on single telomeres derived from Caucasian, Japanese, West-African and Zimbabwean individuals. I have also described a scan for mutations within telomere repeat arrays in pedigrees. The findings described in this chapter have been discussed from the biological standpoints of evolution and mutation of telomeric sequences.

Results

1. TVR-PCR in Caucasian individuals

Unrelated Caucasian DNAs (CEPH and blood DNA) heterozygous at the -30 position flanking the start of the telomere were selected for analysis. These DNAs were subjected to TVR-PCR using allele specific flanking primers (TS-30A/T) and the telomere and variant repeat primers TAG-TelW, TAG-TelX and TAG-TelY (chapter 4). The telomere repeat patterns produced by TVR-PCR were interpreted as telomere repeat codes and recorded in a word processing document. It was possible to code upto 120 repeats in each telomere. Each repeat was assigned a code T=TTAGGG (denoted as - in the figures in this chapter), G=TGAGGG, C=TCAGGG and repeats that did not amplify were coded as N. The first repeat is defined as starting at base +1 as shown in Figure 3.2b. in chapter 3.

A total of 123 telomere maps were obtained from Caucasian individuals (Fig. 5.1a, b, c and d.). Most of the maps were different, but three groups of identical alleles were identified. Two groups were virtually devoid of variant repeats, of these, one group of six alleles started with the motif GTT_n and the other group of 7 alleles started with $GNTT_n$ (Fig. 5.1b. groups 13 and 14). The third group, consisted of just two alleles identical over the 110 repeats coded observed in individuals PeWi and 137501(1) (Fig. 5.1a. group 2).

The telomere maps were grouped according to the haplotypes detected in the DNA flanking the telomere repeat array (chapter 3)(Figures 5.1a, b, c and d.). Subgroups were formed within each haplotypic group. The subgroups were created by comparing the repeats starting at the first repeat with decreasing weight applied to distal repeats. The grouping becomes more subjective when comparing repeats further into the repeat array. A total of 21 different groups containing two or more related alleles were formed, but 13 alleles could not be grouped.

Telomeres associated with haplotype A. Haplotype A has a frequency in Caucasians of 0.49 (Fig. 3.6.). The majority (55/58) of haplotype A associated telomeres start with the motif NNG_{5-8} (Fig. 5.1a.). Six alleles associated with haplotype A were sequenced to determine the nature of the two Ns at the start of the array. The sequence of one of these alleles is described in (Fig. 4.12.). All six alleles started with GAGGG TGAAGG TGAGGG. The TGAAGG repeat cannot be detected by any of the telomere repeat primers, in addition it prevents TAG-TelX from annealing to and priming from the first repeat (GAGGG), therefore the first two repeat units are scored as NN. After approximately 20 repeats, the majority of

Fig. 5.1. TVR-PCR allele databank derived from Caucasian DNAs.

The alleles were derived from DNAs heterozygous at the -30 position, except for alleles with a number in brackets which were derived from DNAs homozygous at the -30 position and the alleles inferred from heterozygous offspring, the number represents the telomere associated with alleles at the DXYS14 locus, for example in the individual 134001 is homozygous for haplotype A, has two telomeres 1 (Fig. 5.1a. group 4) and 2 (Fig. 5.1a. group 5) associated with alleles 1 and 2 at DXYS14.

The alleles were grouped according the haplotype in the flanking DNA to which they are associated. The alleles were then further subgrouped according to similarity, with decreasing weight applied to repeats the further into the telomere repeat array.

G=TGAGGG, C=TCAGGG, '-'=TTAGGG and N=non-amplified repeat.

Fig. 5.1a. Caucasian haplotype A associated Xp:Yp telomere repeat codes

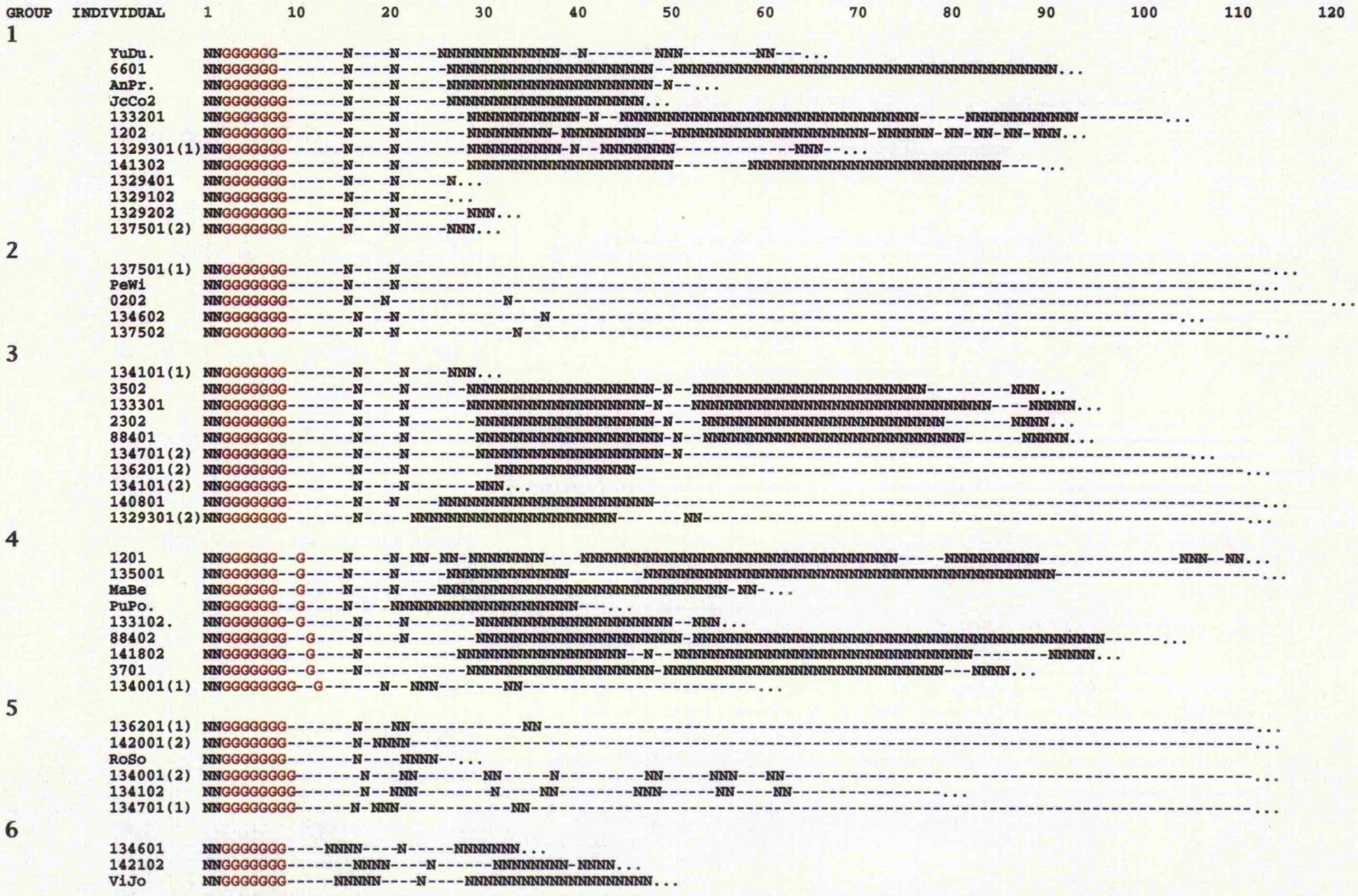


Fig. 5.1a. cont.

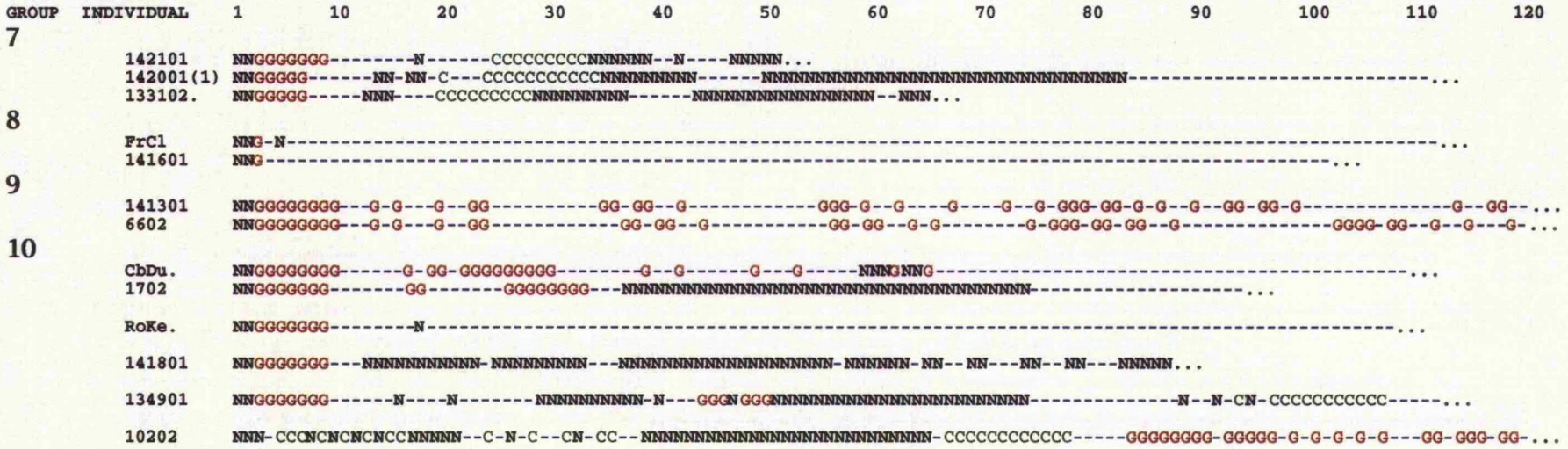


Fig. 5.1b. Caucasian haplotype B associated Xp:Yp telomere repeat codes

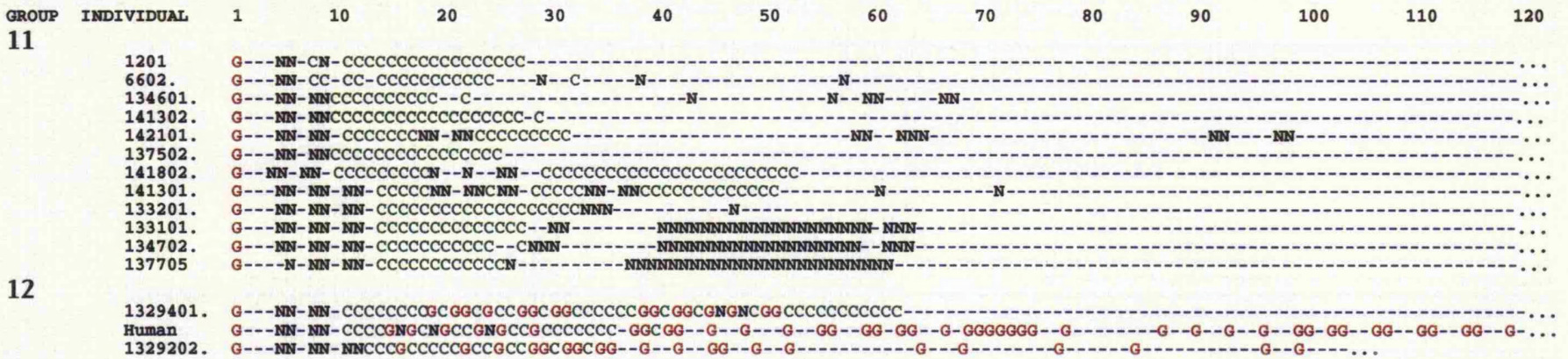


Fig.5.1b. cont.


GROUP INDIVIDUAL 1 10 20 30 40 50 60 70 80 90 100 110 120
13
134002.
135001.
PuPo.
CbDu.
AnPr.
PaBar.
14
1329302.
133301.
88402.
3701.
10402.
10202.
FrCl.
15
JcCo2.
RoSo.
133101.
PeWi.
134702.
141801.
ErBa.
16
10406
140801.
140802.
17
134602.
88401.
136202.
6601.
18
2302.
134102.
MaBe.
19
142002.
142301.
20
3502.
134114.
134901.
1329302
1702.
1329102.

Fig. 5.1c. Caucasian haplotype C associated Xp:Yp telomere repeat codes

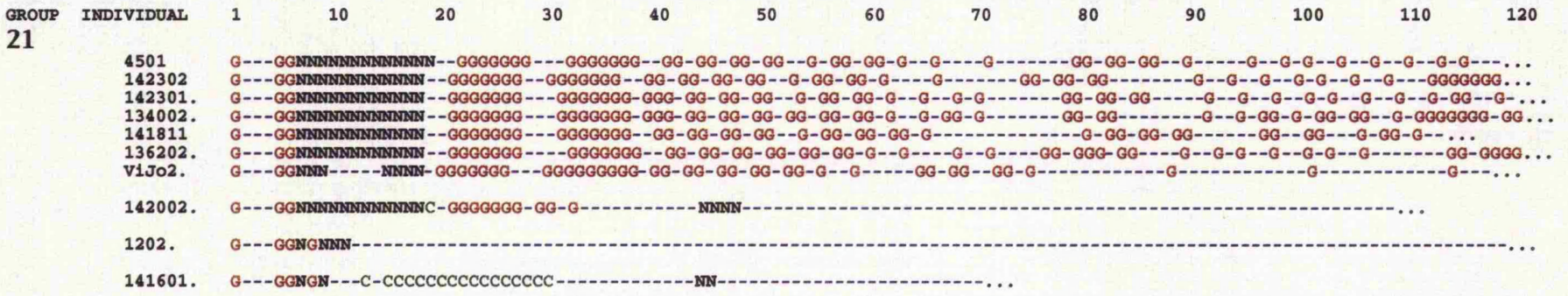
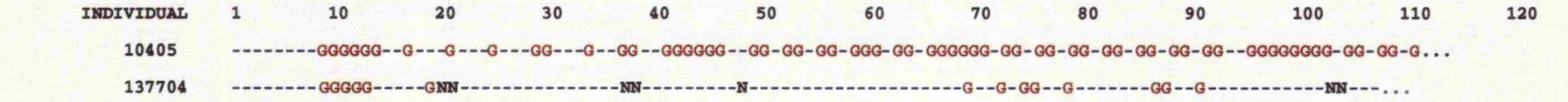


Fig. 5.1d. Caucasian haplotype A´ associated Xp:Yp telomere repeat codes



N.B. Both telomere repeat maps from these A' telomeres start 3 bases distal to the starting point of the all the other telomeres examined, i.e. GAGTTAGGG **instead of** GAGGG....

the telomere codes associated with haplotype A (at least 31 out of 59) contain blocks of eight or more repeats scored as N. I suspect that the majority of these N type repeats may be TTGGGG. This is partly because hybridisation studies showed that this was a common type of telomere repeat unit variant (Allshire *et al.*, 1989), also because a primer designed to detect the TTGGGG repeat type (TAG-TelJ) amplified from a large number of the N-type repeats. However, this primer did not discriminate between different repeat types as well as the other telomere repeat primers, and was not pursued further. Five haplotype A associated telomeres contained TCAGGG repeats in blocks of 12 or less (Fig. 5.1a. group 7, 134901 and 10202) and six contained TGAGGG repeats distal to the first 13 repeat units (Fig. 5.1a. groups 9 and 10, 134901 and 10202). Therefore, the majority of the variation observed in the haplotype A associated telomeres occurs in the distribution of TTAGGG and N-type repeats.

Telomeres associated with haplotype B. Haplotype B has a frequency in Caucasians of 0.44 (Fig. 3.6.). Apart form nine, all haplotype B (n=53) associated alleles start with the motif GT_{1-3} (Fig. 5.1b.). The majority of these telomeres (n=27) contain blocks of at least six TCAGGG repeats most of these are found within the first 50 repeat units. Large blocks of greater than 19 N-type repeats occur in nine telomeres, arranged into two subgroups (11 and 15). Five telomeres contain TGAGGG repeats distal to the first repeat unit mostly interspersed with TTAGGG repeats. As stated above, there are two groups of alleles virtually devoid of variant repeat types (Fig. 5.1b. groups 13 and 14). These telomeres are not informative, therefore, it is difficult to determine whether the telomeres within these groups are derived from a common ancestor. The lack of telomere variant repeats in these telomeres could reflect the recent action of telomerase extending a shortened telomere, either *in vivo* or during the formation and culturing of the lymphoblastoid cell line from which the CEPH DNAs are derived. The telomere repeat codes obtained from the haplotype B associated telomeres are easier to read (using the flanking primer TS-30A) than those associated with the haplotype A (using the flanking primer TS-30T). This is primarily because of the large number of N-type repeats in telomeres associated with haplotype A, which can make coding difficult. Therefore, codes extending further into the telomere repeat array have been obtained from haplotype B associated telomeres, the majority of these (n=50) are composed solely of TTAGGG repeats beyond 110 repeats, only three of these telomeres contain informative variation extending beyond this point (for example, 1702 Fig. 5.1b.). The variation in these telomeres consists of TGAGGG repeats interspersed with TTAGGG repeats and inspection of high molecular bands in the TVR-PCR gels shows that this variation appears to extend for a large number of repeats distally to that resolved in TVR-PCR gels (130 repeats plus).

Telomeres associated with haplotype C. Haplotype C has a frequency in Caucasians of 0.05 (Fig. 3.6.). All (n=10) haplotype C associated alleles start with the motif GTTTGGN (Fig. 5.1c.). Seven of these telomeres can be put into one group consisting of a complex interspersion pattern of TGAGGG and TTAGGG repeats. Informative variation in this group of telomeres extends beyond 120 repeats. One haplotype C associated telomere contains a block of TCAGGG repeats (Fig. 5.1c. 141601), this telomere starts with the GTTTGGGN sequence characteristic of the other haplotype C associated telomeres.

Telomeres associated with haplotype A´. Haplotype A´ is rare in Caucasians with a frequency of 0.02 (Fig. 3.6.), it is the same as haplotype A at the -842, -652, -427 and -414 polymorphic positions but at the -30 and -13 positions it switches to the sequence found in haplotypes B or C (see chapter 3). The telomere maps from two A´ Caucasian telomeres were obtained. These telomeres are different from any previously mapped telomere (Fig. 5.1d.). They are the only telomeres among those analysed that start with TTAGGG repeats, also the TVR-PCR maps do not start at the same position as the other maps. Bands cut from the dried TVR gels were sequenced from both these A´ telomeres to determine exactly were the telomere repeat array started. The TVR-PCR maps started three bases distal (at +3) compared to all the other telomeres mapped. The first repeat, which is usually GAGGG, had been truncated to GAG. TTAGGG repeats start at this position giving the sequence GAG TTAGGG TTAGGG.

2. TVR-PCR in Japanese individuals

DNA from unrelated Japanese individuals heterozygous for the -30 flanking polymorphism were subjected to TVR-PCR, and coded in the same way as for the Caucasian DNAs.

A total of 63 telomere maps were obtained, all were different again with the exception of a group of eight telomeres associated with the flanking haplotype B devoid of telomere repeats. Four pairs of telomeres were identical over the first 50 repeats (Fig. 5.2a.). Inspection of the top of the TVR-PCR autoradiogram of these telomeres revealed further variation but it was not possible to interpret this as a code. No haplotype C or A´ associated telomeres were obtained since these are very rare in the Japanese population. A total of 5 groups of related alleles were formed with 16 alleles that could not be grouped.

Telomeres associated with haplotype A. In the Japanese population haplotype A has a frequency of 0.31 (Fig. 3.6.). A total of 30 telomere maps

Fig. 5.2. TVR-PCR allele databank derived from Japanese DNAs.

The alleles were derived from DNAs heterozygous at the -30 position. The alleles were grouped according the haplotype in the flanking DNA with which they are associated. The alleles were then further subgrouped according to similarity, with decreasing weight applied to repeats further into the telomere repeat array.

G=TGAGGG, C=TCAGGG, '-'=TTAGGG and N=non-amplifiable repeat.

Fig. 5.2a. Japanese haplotype A associated Xp:Yp telomere repeat codes

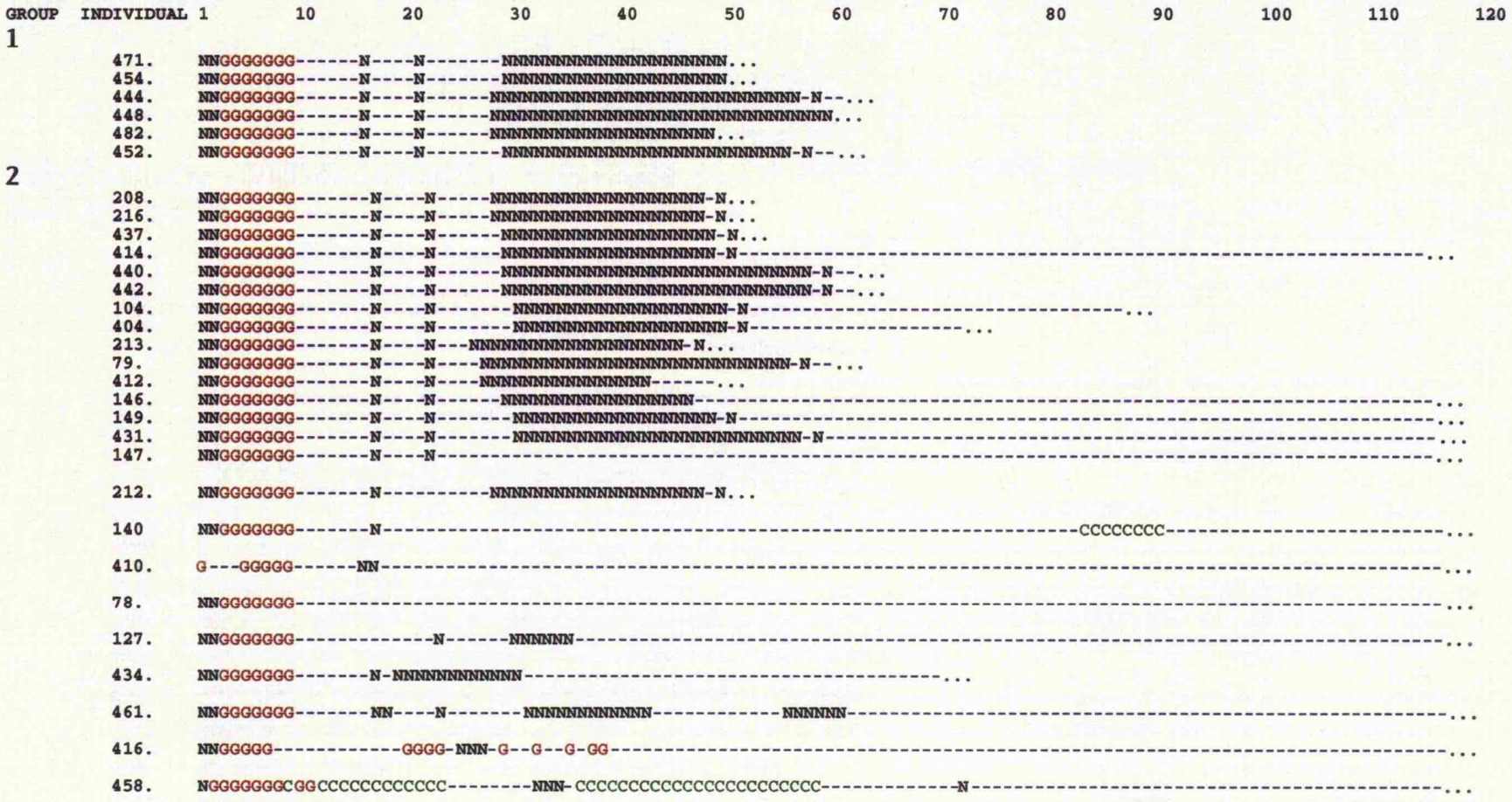
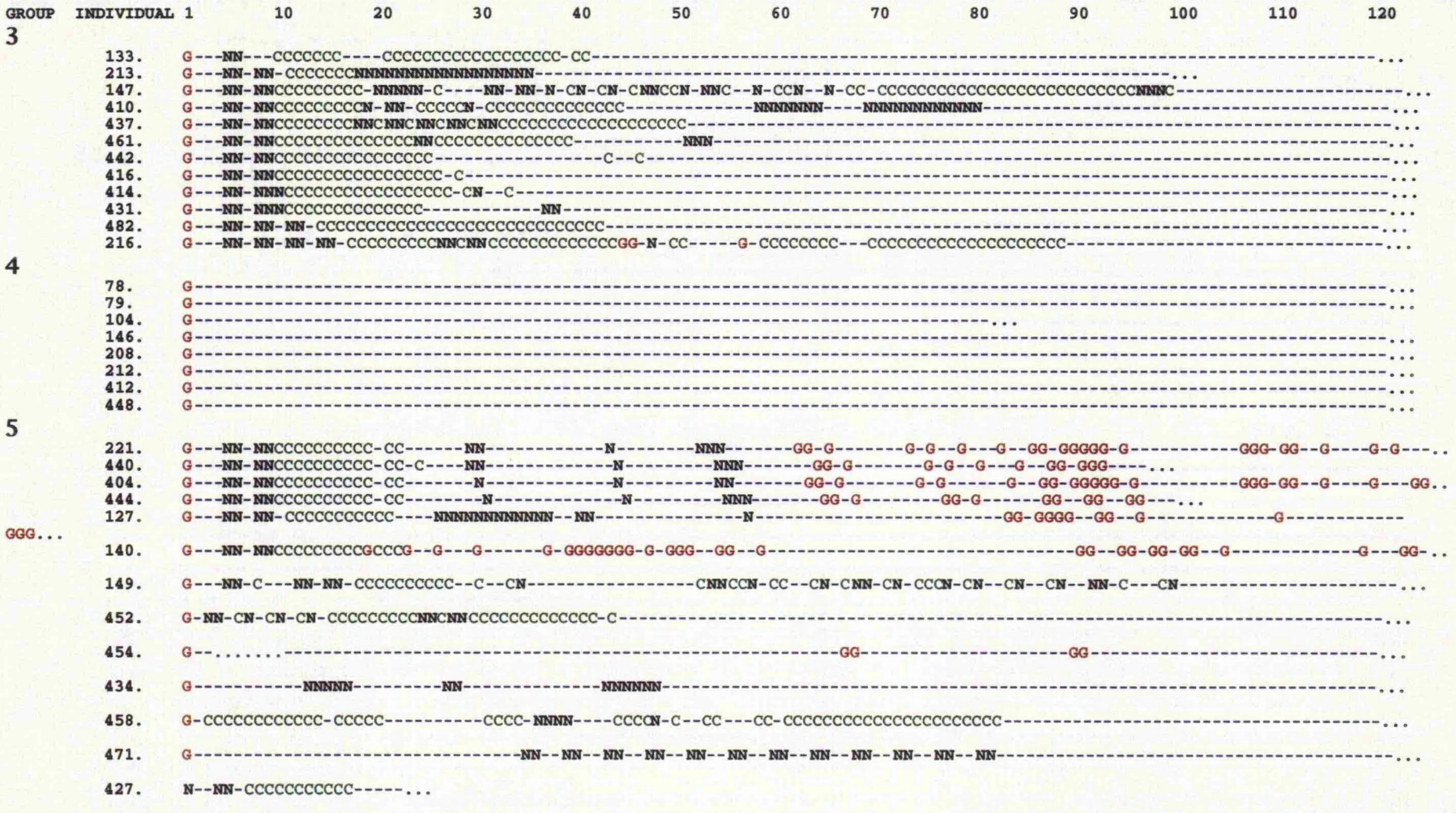


Fig. 5.2b. Japanese haplotype B associated Xp:Yp telomere repeat codes



associated with the flanking haplotype A were obtained (Fig. 5.2a.). These telomeres are very similar to the haplotype A alleles observed in Caucasians (Fig. 5.1a). Again the majority (n=28) start with the sequence NNG_{5-7}. Also, the majority of these telomeres contain N-type repeats in a block beginning 25-30 repeats into the telomere. Only two telomeres contain TCAGGG and just one telomere contains TGAGGG repeats distal to the first nine repeat units.

Telomeres associated with haplotype B. In the Japanese population haplotype B has a frequency of 0.66 (Fig. 3.6.). A total of 33 telomere maps were obtained from telomeres associated with haplotype B (Fig. 5.2b.). All apart from one telomere started with the sequence GT_{1-3}. The majority (n=22) contain one or more blocks of TCAGGG repeats. Eight telomeres contain TGAGGG repeats six of which are contained in the same subgroup (Fig. 5.2b. group 5). As stated above, there is a large group of eight telomeres devoid of telomere variant repeats (Fig. 5.2.b. group 4). Again there were telomeres which had informative variation extending beyond the resolution of the electrophoretic systems used (Fig. 5.2b. group 5).

3. TVR-PCR in West-African and Zimbabwean individuals

DNA from West-African (Nigerian) and Zimbabwean individuals heterozygous for the -30 flanking polymorphism were subjected to TVR-PCR. A total of 43 telomere maps were obtained. Again they were all different except for five telomeres devoid of telomere variant repeats, one associated with haplotype B and four with haplotype A′. The telomere repeat codes from these populations were pooled and grouped according to flanking haplotype then subgrouped in the same way as for the Caucasians. A total of nine groups of two or more related alleles were formed with 12 alleles that could not be grouped.

Telomeres associated with haplotype A. In the West-African population haplotype A has a frequency of 0.5 (Fig. 3.6.). A total of 22 haplotype A associated telomeres were obtained (Fig. 5.3a.). The distribution of telomere and variant repeats were somewhat different from the haplotype A associated telomeres observed in the Caucasian and Japanese populations. Fifteen telomeres all started with the sequence $GNNTCCN_{8-24}$ (Fig. 5.3a. groups 1, 2 and 3, plus single alleles 1, AF13 and AF8) and of these, groups 1 and 3 and allele 1 contained blocks of 9-28 TCAGGG repeats. Six telomeres started with the sequence NNG_{6-7} and contained large blocks of N-type repeats (Fig. 5.3a. groups 4 and 5), as observed in the Caucasian and Japanese populations. The largest group of telomeres associated with haplotype A, a group consisting of seven telomeres (Fig. 5.3a. group 1), are very similar to one telomere observed in Caucasians (10202 Fig. 5.1a.). This large

Fig. 5.3. Telomere maps obtained from West-African and Zimbabwean individuals.

Telomere maps from West-African individuals can be identified by the prefix AF on the lab identifier and the remaining alleles are derived from Zimbabwean individuals. The alleles were derived from DNAs heterozygous at the -30 position. The alleles were grouped according the haplotype in the flanking DNA with which they are associated. The alleles were then further subgrouped according to similarity, with decreasing weight applied to repeats further into the telomere repeat array.

G=TGAGGG, C=TCAGGG, '-'=TTAGGG and N=non-amplifiable repeat.

Fig. 5.3a. West-African and Zimbabewean haplotype A associated Xp:Yp telomere repeat codes

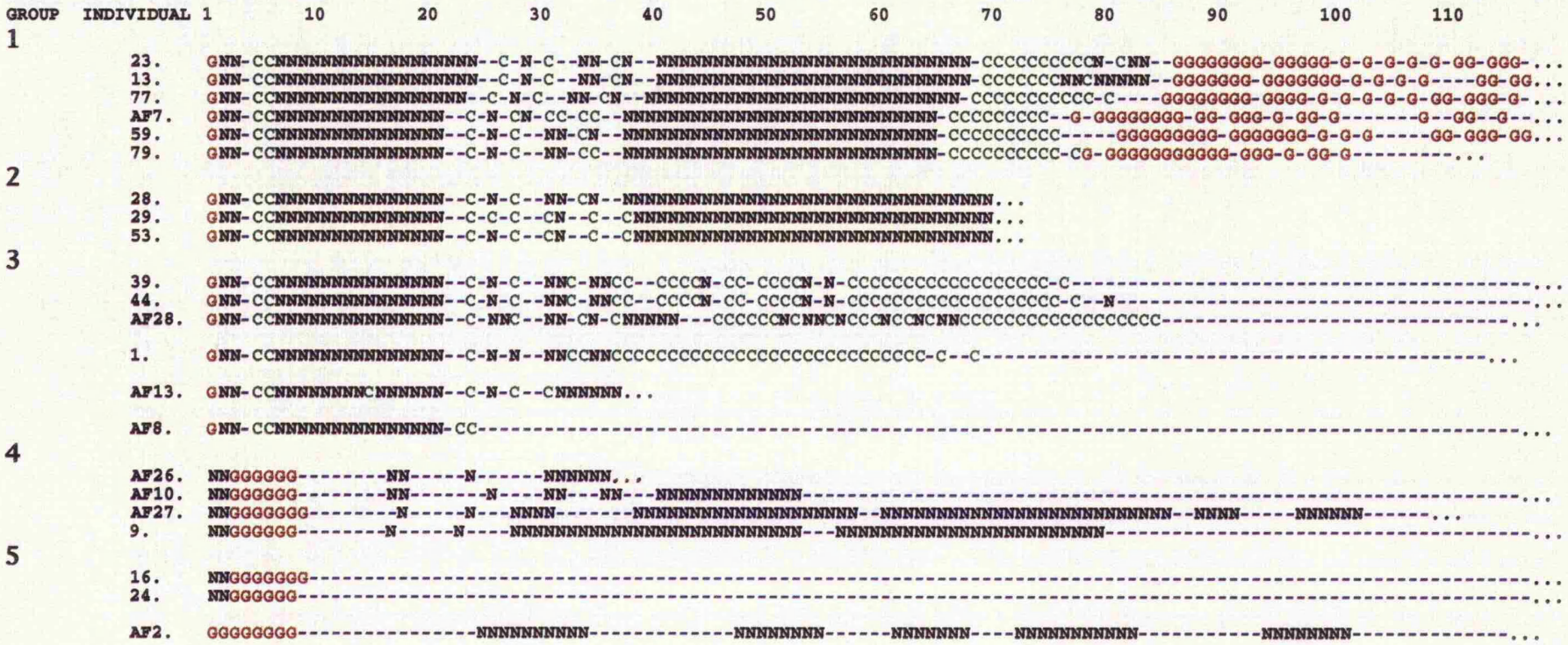


Fig. 5.3b. West-African and Zimbabewean haplotype B associated Xp:Yp telomere repeat codes

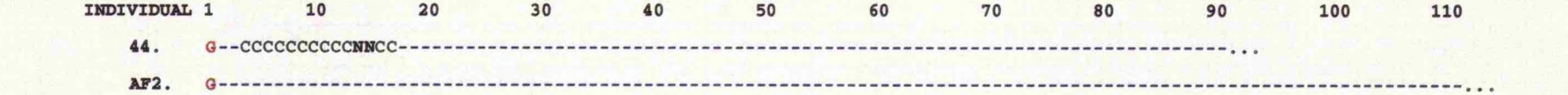


Fig. 5.3c. West-African and Zimbabewean haplotype C associated Xp:Yp telomere repeat codes

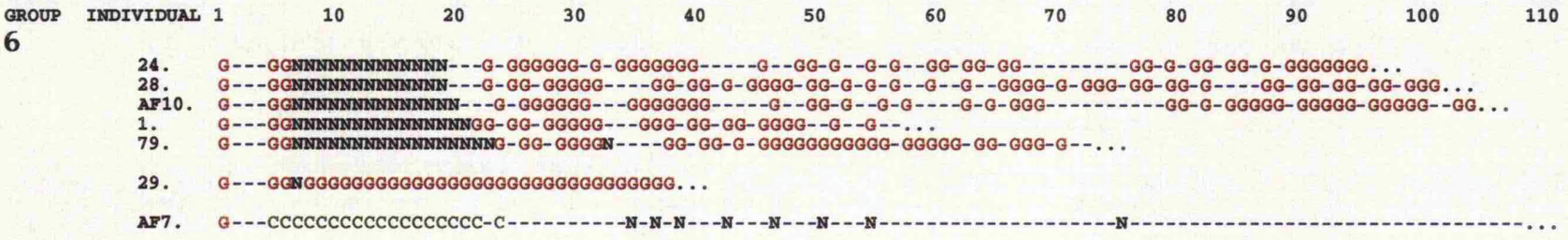
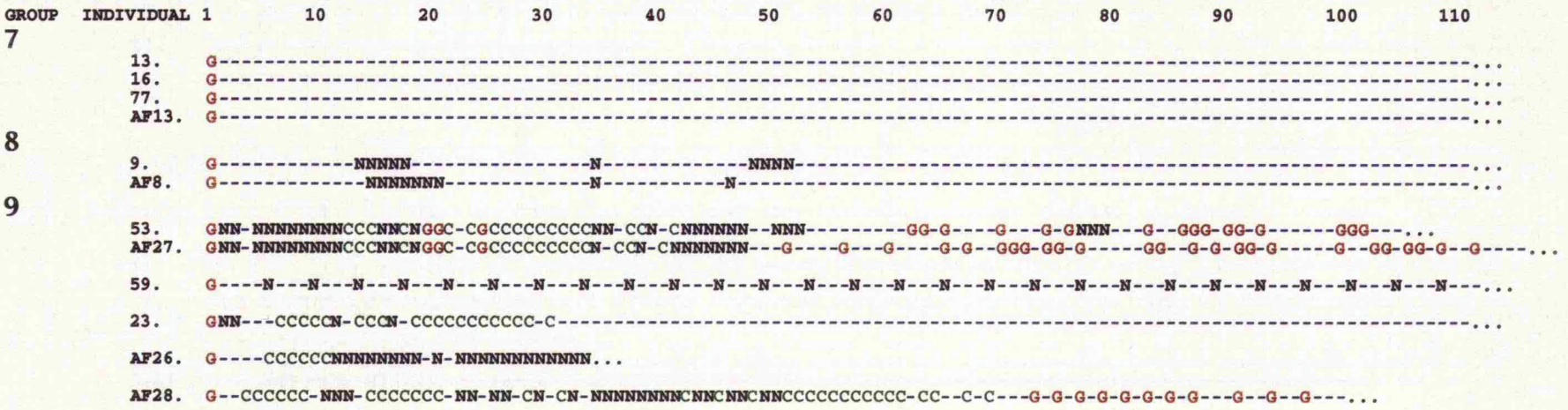


Fig. 5.3d. West-African and Zimbabewean haplotype A´ associated Xp:Yp telomere repeat codes



group of related telomeres have informative variation extending beyond the resolution of the electrophoretic systems used.

Telomeres associated with haplotype B. Only two examples of haplotype B associated telomeres were obtained because this haplotype is rare in these populations (with a frequency of 0.02 in the West-African population see chapter 3, Fig. 3.6.). One is devoid of telomere variant repeats the other has a block of TCAGGG repeats (Fig. 5.3b.). Both of these telomeres are similar to those found associated with haplotype B in the Caucasian and Japanese telomeres.

Telomeres associated with haplotype C. In the West-African population haplotype C has a frequency of 0.24 (Fig. 3.6.). Seven haplotype C associated telomere maps were obtained (Fig. 5.3c.). All apart from one start with the sequence $GTTTGGN_{1-17}$, the same as that observed in the Caucasian haplotype C associated telomeres. Five telomeres form one group of related telomeres (Fig. 5.3c. group 6) similar to the large group observed in Caucasians. One telomere contains a large block of TCAGGG repeats similar to that found in Caucasian telomeres associated with haplotype B. One very unusual telomere (observed in Zimbabwean individual 29, Figure 5.3c.) appears to consist entirely of TGAGGG repeats over the first 100 repeat units, it was not possible to define the distal end of these variant repeats.

Telomeres associated with haplotype A′. Haplotype A′ is fairly common in these populations (with a frequency of 0.24 in the West-African population see chapter 3, Fig. 3.6.). This haplotype is the same as haplotype A over the flanking polymorphisms -842, -652, -427 and -414 but at the -30 polymorphism it is the same as haplotypes B or C (the -13 polymorphism has not been assayed in these populations). A total of 12 haplotype A′ associated telomere maps were obtained (Fig. 5.3d.). These telomere maps have some features in common with the two haplotype B telomeres observed in these populations and to the haplotype B telomeres observed in Caucasians and Japanese. All apart from three start with the sequence GTT, four of these telomeres are devoid of telomere variant repeats (like the telomeres associated with haplotype B in Caucasians, Japanese and Africans) and five have blocks of TCAGGG repeats. The alleles of group 9 do not bear much resemblance to haplotype B telomeres and only appear to be associated with haplotype A′. There is one very unusual telomere observed in the Zimbabwean individual 59. This telomere has a higher order structure consisting of the motif NTTT tandemly repeated, the distal most end of this higher order structure is not defined. These haplotype A′ associated telomeres are very different to the two Caucasian haplotype A′ associated telomeres (Fig. 5.1d.).

4. Allele alignments

Groups of related alleles were assembled as shown in figures 5.1., 5.2. and 5.3.. The alignments observed within these groups can be improved by the inclusion of gaps. Groups of alleles with informative patterns of repeat unit interspersion are shown in figure 5.4. The majority of alignments can be improved by the addition of gaps into large blocks of like repeats. This reflects expansion or contractions that must have occurred during the evolution of these lineages. The best alignments occur approximately within the proximal 90 repeat units, these alignments then deteriorate in the distal regions. The majority of the groups shown in figure 5.4. contain alleles from one racial group. Although most of the alleles from any one haplotypic group share sequence motifs between races particularly at the beginning of the telomere repeat array, there are differences further in to the array which make it difficult to create alignments. This is reflected in the aligned group 8 (Fig. 5.4.), where it is possible to align these alleles in Caucasians, but it is difficult to include the African alleles (shown below the Caucasian alleles). In this group it would be possible to align the African alleles with the Caucasian alleles over the first 30 repeats, but distal to this the differences are too great to create alignments. There are examples of alignments between racial groups, such as those shown in group 9 (Fig. 5.4.). These are very similar alleles, but have a simple structure therefore the opportunity for informative variation in these alleles is reduced.

5. Heterozygosity

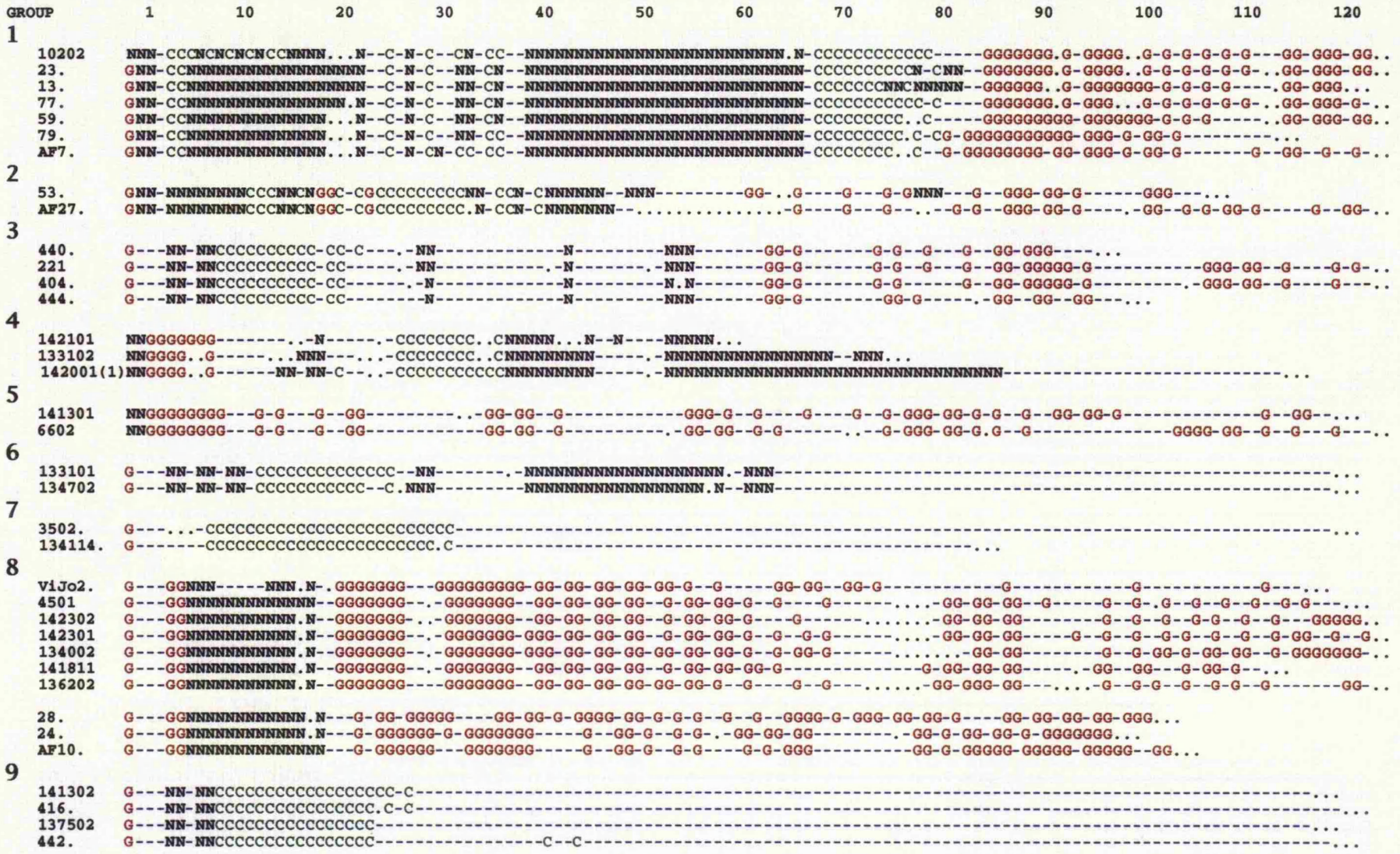
Allelic diversity, heterozygosity and mutation rate at the Xp:Yp telomere can be estimated using the population data. The allelic diversity (θ) was determined under the infinite-allele model assuming selective neutrality, $n_a = (\sum_{i1}^{2i} \frac{\theta}{\theta+i-1})$ where n_a is the number of different alleles observed in a sample of i alleles (Ewens, 1972) using a computer programme (written by A. J. Jeffreys in Microsoft Quickbasic^{TM}). In a sample of 123 Caucasian alleles 111 different alleles were observed, this gives an allelic diversity of 506 which gives an estimated heterozygosity of 99.8% [100-(100/506)]. Solving the equation $\theta=4N_e\mu$, where N_e is the effective population size which in humans has been estimated at 10^4 (Morton, 1982), and μ is the mutation rate, allows the mutation rate to be estimated at 1.27×10^{-2} ($\mu=(506/4)/10^4$). In Japanese the allelic diversity and heterozygosity is estimated to be 211, with a heterozygosity of 99.5.%. In West-Africans $\theta=110$ and the heterozygosity is estimated at 99.1%, in Zimbabweans $\theta=108$ and the heterozygosity is estimated at 99.1%.

Fig. 5.4. Aligned telomere repeat alleles.

The inclusion of a period '.' in a repeat code represents the inclusion of a gap to improve alignments. The majority of the groups shown are from one population and all the alleles within a group are associated with the same haplotype in the flanking DNA. Group 1 = African, hap. A and one Caucasian, hap. A (10202); group 2 = African, hap. A´; group 3 = Japanese, hap B; groups 4 and 5 = Caucasian, hap A; groups 6 and 7 = Caucasian, hap B; group 8 Caucasian (top) and African (bottom), hap C; group 9 = Caucasian and Japanese (416 and 442), hap B.

G=TGAGGG, C=TCAGGG, '-'=TTAGGG and N=non-amplifiable repeat.

Fig. 5.4. Groups of related alleles with the inclusion of gaps to improve alignments

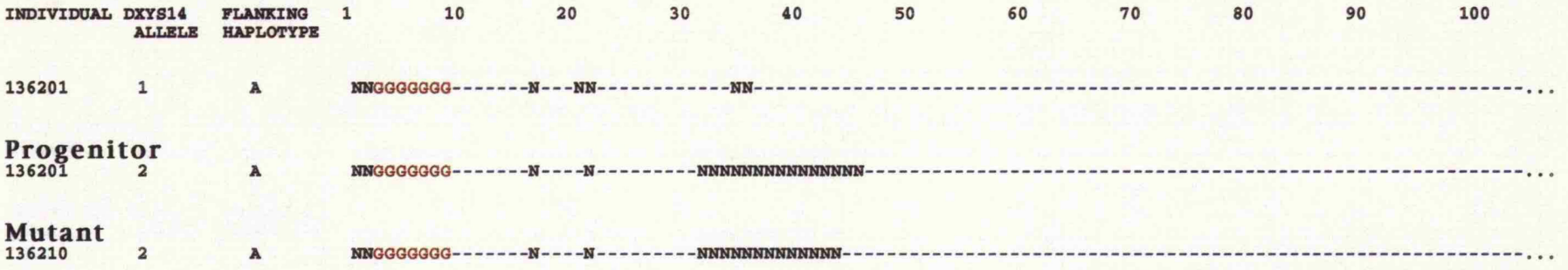


Chapter 5. Figure 4

6. Mutation in the Xp:Yp telomere

In order to investigate further the mutational mechanisms generating the variation observed in the telomere and variant repeat unit distribution in the Xp:Yp telomere, a scan for mutation was carried out through pedigrees (CEPH). This was carried out using segregation data of the polymorphisms in the flanking DNA already obtained during segregation analysis with DXYS14 (chapter 3). DNA from heterozygous individuals of the same family that should contain the same telomere from either one of the parents, were combined into pools of no more than three DNAs. The pools of DNA were subjected to TVR-PCR, any differences detected were examined by TVR-PCR on each member of the pool in which the difference was detected. Detection of differences was made easier by running the reactions for detection of the same repeat type in adjacent lanes of the polyacrylamide gel. Using this technique a total of 160 meioses were examined. A single putative mutation event was detected in the CEPH pedigree 1362, this occurred in a male meiosis in the father 136201 and was transmitted to his daughter 136210. The father was homozygous for haplotype A, both of the telomere alleles were deduced from the analysis of his heterozygous offspring. Since the haplotypes in the DNA flanking the telomere are uninformative, information as to which allele contained the mutation was derived from the segregation data in the CEPH database (version 60) for DXYS14 (Fig. 5.5.). One mutation in 160 meioses represents a mutation rate of 6.25×10^{-3} (with 95% upper and lower confidence limits of 2.93×10^{-2} and 3.21×10^{-4} respectively). This rate is in good agreement with the mutation rate estimated above. The progenitor and the mutant allele are shown in figure 5.5., the mutation consists of an apparent two repeat unit contraction of a block of N-type repeats (Fig. 5.5.). However, it must be taken into consideration that the DNAs used during this screen for mutation are obtained from lymphoblastoid cell lines. In such cell lines somatic mutations have been detected which could be mistaken as germline mutations (Royle *et al.*, 1993; Banchs *et al.*, 1994). To confirm that mutations in the CEPH pedigrees are genuine, one must observe the stable transmission of the mutant allele through a pedigree. However, no offspring from the individual in which the mutation has been detected are available.

Fig. 5.5. A mutation detected in telomere and variant repeat unit distribution in the Xp:Yp telomere



Discussion

Allelic variation. The analysis of the distribution of telomere and variant repeat units in the proximal 720bp (120 repeats) of the Xp:Yp telomere repeat array has revealed high levels of allelic variation. This variation resulted in heterozygosities in excess of 99%, comparable to the most hypervariable minisatellite loci (Wong *et al.*, 1986). The underlying cause of this diversity must be a high mutation rate in the order of 1.27×10^{-2} (estimated from the allelic diversity of Caucasians and the human effective population size).

Repeat unit distribution. The majority of TGAGGG repeats occurred in blocks of nine or less repeats, and generally occurred in blocks of one or two repeats interspersed by TTAGGG repeats. TGAGGG repeats were often observed in telomere repeat arrays beyond the limit of good resolution with the electrophoretic systems used. The TCAGGG repeats usually occurred in large blocks of six repeats or more, some of these blocks were quite large for example a 28 repeat unit block observed in Zimbabwean individual 1 (Fig. 5.3a.). All TCAGGG repeats detected occurred within the proximal 105 repeats. The observed variation in repeat type distribution, suggests that the evolution of the types of repeats assayed cannot be influenced by any functional constraints. These variant repeat units probably have no function, since if they were functionally constrained it is likely that the distribution would not vary. However, there may be some functional constraint acting on telomere variant repeat units to prevent the complete degeneration away from a 6bp repeat with a run of three G-residues in proximal most regions of the telomere repeat array. This is reflected by the fact that the GGG portion and 6bp periodicity of the variant repeat units assayed appears to remain constant. Such functional constraints may prevent the spread of mutations in the GGG portion but would allow spread of mutations in the TTA portion of the repeat unit. Perhaps a disruption in the GGG affects the affinity for telomere binding proteins or it may affect the chromatin characteristics of the telomere. The mammalian telomere binding protein TRF has a high binding specificity for TTAGGG and does not bind well to variant repeats such as TTGGGG (Zhong *et al.*, 1992; Chong et al., 1995). Telomere junction fragments isolated from pigs and chickens revealed that the proximal regions of the telomeres are highly degraded such that it is often difficult to define the start of the telomere repeat array (E. Scanlan and P. Thompson pers. comm.). It is possible that the human Xp:Yp telomere is relatively young and the apparent conservation of the general telomere repeat structure is because the telomere has not had time to degrade in the same way as that observed in pigs and chickens. In the haplotype A associated telomeres the distal most end of the variant repeats was not defined, but in the haplotype B, C and A´ associated telomeres the majority of the variation was

observed within the first 120 repeat units. Variation past the 120 repeat unit limit cannot be assayed with this technique, however, it has been observed by Southern blot hybridisation analysis, that the distal limit of the variant repeat TGAGGG is 1.9kb and 2.0kb for most of the TTGGGG repeats (Allshire *et al.*, 1989). The proximal location of variant repeat units in the telomere repeat array could reflect functional constraints preventing the spread of these repeats throughout the telomere, which could compromise telomere function in a similar manner to that observed upon the introduction of variant repeats to the terminus of *Tetrahymena* telomeres (Yu *et al.*, 1990; Yu and Blackburn, 1991). The proximal distribution of telomere variant repeats could also be a consequence of the action of telomerase repairing broken or shortened telomeres, resulting in the replacement of variant repeats with the canonical telomere repeat unit.

The sequence of the West-African and Caucasian haplotype A differs at the intervening polymorphisms to give a sequence divergence of 0.73% (see chapter 3, Table 3.7.). Analysis of telomeres associated with these haplotypes revealed that the majority of Caucasian haplotype A associated telomeres share similarities, particularly the presence of the motif NNG_{5-8} and the presence of large blocks of N-type repeats (Fig. 5.1a.). The haplotype A associated telomeres derived from the African populations appeared to have two classes of similar telomeres, one class with motifs also characteristic of Caucasian haplotype A (NNG_{5-8}) associated telomeres and a second class which start with the motif $GNNTCCN_{8-24}$ (Fig. 5.3a.). The sequence analysis of the flanking DNA was carried out on homozygous individuals and therefore was not specifically of the flanking DNA of the telomeres to which TVR-PCR has been applied. However, given the observed evolution along haploid lineages, it seems likely that the African telomeres with the motif $GNNTCCN_{8-24}$ may be associated with the African haplotype A sequence, and those African and Caucasian telomeres with the NNG_{5-8} motif may be associated with the Caucasian haplotype A sequence.

Mechanisms generating variation. The telomere repeat codes generated by TVR-PCR can be arranged into groups according to similarity. The alleles within these groups have the same flanking haplotype. This suggests that the evolution of these sequences must be largely along haploid lineages and means that the mutational mechanisms underlying the observed variation must predominantly be of an intra-allelic nature. Such intra-allelic mechanisms may include replication slippage (Levinson and Gutman, 1987; Murphy *et al.*, 1989; Schlotterer and Tautz, 1992) and sister-chromatid exchange (Schlotterer and Tautz, 1994), but these two mechanisms cannot formally be distinguished. The majority of differences between alleles within a group of related alleles can be accounted for by small localised expansions or contractions within blocks of identical repeat

units. The variation observed in these telomere repeat codes appears to increase in the more distal regions of the mapped arrays, whereas, the most proximal repeats appear to be in strong linkage disequilibrium with the haplotypes observed in the flanking DNA. This data is consistent with a replication slippage type of mechanism in which it is proposed that the longer the run of identical repeat units the greater the opportunity for replication slippage to occur (Weber, 1990; Richards and Sutherland, 1994; Eichler *et al.*, 1994). The reduced variation in the proximal part of the telomere repeat array could result from of an anchoring effect of the DNA flanking the telomere repeat array, so reducing the opportunity for replication slippage. In addition, the strong linkage disequilibrium in the DNA flanking the telomere repeat array (chapter 3) suggests that there may be reduced levels of recombination in this region. This may result in a prevalence of mutations by intra-molecular mechanisms during the evolution of the Xp:Yp telomere and may preclude the involvement of unequal sister-chromatid exchange which is an interaction involving exchanges between molecules.

One particularly interesting telomere has been observed in the Zimbabwean individual 29. This telomere appears to consist entirely of TGAGGG repeats at least within the proximal 120 repeat units of the telomere repeat array. This may be an example of an expansion of a large block of TGAGGG repeats such that the block of repeats expanded to size such that a complete Okazaki fragment could be contained within the block (Richards and Sutherland, 1994; Eichler *et al.*, 1994). Human Okazaki fragments are estimated to be between 150 to 200nt long (Thommes and Hubscher, 1990), which could mean that blocks of identical telomere repeats greater than 25 to 33 repeats long could be prone to large replication slippage events. This may result in a very large expansion of repeats, analogous to that observed at disease loci in which triplet repeat expansion has been implicated (Yu *et al.*, 1991; Brook *et al.*, 1992).

Extensive higher order structure was observed in the haplotype A´ associated telomere of Zimbabwean individual 59 (Fig. 5.3d), and in others. How this is generated is not known. It may include unequal exchanges similar to those proposed for the evolution of the higher order structures observed in α-satellite DNA (Greig *et al.*, 1993; Warburton *et al.*, 1993). However, the large differences in repeat unit length and complexity between telomere repeats and α-satellite (6bp to 171bp) may make such comparisons inappropriate.

The association of groups of related alleles with the flanking haplotype is not absolute. There appear to be examples of repeat unit motifs that are shared between haplotypic groups. For example, the Caucasian 141601 haplotype C

associated telomere (Fig. 5.1c.). This telomere starts with the motif GTTTGN common to virtually all haplotype C associated telomeres, but contains a block of 16 TCAGGG repeats common in Caucasian haplotype B associated telomeres. This could be an example of an allele that has undergone an inter-allelic exchange event either by recombination or a gene conversion-like process. However, if such inter-allelic events occur they must play a minor role in the evolution of the Xp:Yp telomere compared to that of intra-allelic mechanisms.

Mutation in the Xp:Yp telomere. The mutation rate at the Xp:Yp telomere as been measured directly at 6.25×10^{-3} ($\pm2.93\times10^{-2}$ and 3.21×10^{-4}) per gamete. One mutation was detected which resulted in the loss of two repeat units in a block of 15 N-type repeats. It was not possible to determine whether this mutation was caused by inter or intra-allelic mechanisms since the other allele in the father was not informative (Fig. 5.5.). However, this mutation is indicative of the types of events that would be predicted from the comparison of related alleles and therefore may be a result of an intra-allelic replication slippage event.

Haplotype A´ associated telomeres. The complete sequence of A´ flanking haplotypes has not been determined in Caucasians or Africans (see chapter 3). The telomere maps derived from two Caucasian haplotype A´ associated telomeres are different to those derived from 12 African telomeres (Fig. 5.1d. and 5.3d.). This suggests that these haplotypes are not derived from a common ancestor. However, the two A´ telomeres in Caucasians share common features including a 3bp polymorphism in the start site of the telomere and could therefore be derived from a common ancestor. The Caucasian haplotype A´ may have arisen by a complex mutational event which somehow involved the truncation of the first repeat of the telomere repeat array from GAGGG to GAG, possibly coupled with an event that converted the flanking polymorphisms -30 and -13 from one haplotype to another.

The African haplotype A´ telomere codes share common features with haplotype B associated telomere codes derived from Africans, Caucasians and Japanese. The -30 flanking polymorphism is the same as haplotype B but the sequence at the -842, -652, -427 and -414 polymorphisms is the same as haplotype A. Therefore, it is possible that the A´ haplotype in Africans is derived from exchange events in the flanking DNA covering the -30 and -414 polymorphisms between haplotypes A and B.

Length polymorphism at the ends of human chromosomes. During the course of the analysis of variation in the Xp:Yp telomere and flanking DNA, it has become evident that there is a small amount of variation in the start site of the

telomere repeat array. This is observed in the Caucasian A´ telomeres, where the start site of the telomere is 3bp distal to all the other telomeres analysed. This polymorphism must be rare, with a maximum frequency equal to that of the A´ haplotype (0.02, see chapter 3, Fig. 3.6.). However, the variation in the start of the Caucasian haplotype A´ associated telomeres is not on the same scale as those observed at other chromosome ends. Large length variants have been observed at the 16p telomere of upto 260kb (Wilkie *et al.*, 1991; Harris and Thomas, 1992), the 2q telomere (Macina *et al.*, 1994) and may be present at the telomeres of other chromosomes (Ijdo *et al.*, 1992). The polymorphisms in the DNA adjacent to the Xp:Yp telomere repeat array are in Hardy-Weinberg equilibrium (Chapter 3) and the telomere repeat array can be amplified from both alleles of all the heterozygous individuals. This suggests that the length polymorphism such as that observed at 16p does not occur within 2kb of the start of the Xp:Yp telomere. However, this does not preclude the possibility of such a polymorphism proximal to this locus, but physical maps of the pseudoautosomal region which include DXYS14 show no evidence of this (Brown, 1988). Length polymorphisms detected at other chromosome ends are thought to result from subterminal terminal exchanges (Wilkie *et al.*, 1991; Ijdo *et al.*, 1992; Macina *et al.*, 1994). The lack of such length polymorphism at the Xp:Yp telomere is probably not surprising given the absence of human subterminal sequence families in this region (Brown *et al.*, 1990).

Linkage disequilibrium. Strong linkage disequilibrium has been observed in the DNA flanking the Xp:Yp telomere (chapter 3). The analysis of telomere repeat codes associated with these haplotypes showed that the linkage disequilibrium does extend into the telomere repeat array, as telomeres from each haplotypic group begins with a characteristic motif. For example, all (n=17) haplotype C associated telomeres derived from Caucasians and Africans start with the motif GTTTGGN with the exception of one which starts GTTTCCC. However, the mutational mechanisms that generate variation in these sequences appear to break down this linkage disequilibrium within the first few repeats of the telomere repeat array.

Conclusions

TVR-PCR has revealed an extremely high degree of allelic variation in the distribution of telomere and variant repeats in the proximal 120 repeats of the Xp:Yp telomere. This results in heterozygosities in excess of 99% in all the populations analysed. The telomere repeat codes can be grouped with similar alleles associated with the same haplotype in the flanking DNA. This suggests

that the predominant mechanisms involved in the evolution of these sequences must be of an intra-molecular nature, most likely involving replication slippage although unequal sister-chromatid exchange cannot be ruled out. The observed variation must be due to a high underlying mutation rate, this has been directly measured at 6.25×10^{-3} ($\pm2.93\times10^{-2}$ and 3.21×10^{-4}) per gamete.

Chapter 6

ANALYSIS OF SEQUENCES ORTHOLOGOUS TO THE DNA FLANKING THE HUMAN Xp:Yp TELOMERE IN NON-HUMAN PRIMATES

Summary

A high frequency of sequence polymorphism and strong linkage disequilibrium has been detected in the DNA flanking the human Xp:Yp telomere creating a limited number of highly diverged haplotypes (chapter 3). Analysis of chimpanzees (*Pan troglodytes*) and gorillas (*Gorilla gorilla*) revealed a high degree of sequence divergence from the human haplotype A (3.5% in chimpanzees and 4.3% in gorillas). These estimates of divergence are higher between these species than divergence estimates derived from other regions of the genome. However, the chimpanzee and gorilla sequences were more similar to the human haplotype A at the polymorphic positions that define this haplotype in humans than to the other haplotypes B and C (chapter 3). Therefore, the genome of the ancestral species that gave rise to the human, chimpanzee and gorilla lineages may have contained a haplotype A-like sequence. No sequence polymorphism was detected in either chimpanzees or gorillas, however only a small number of individuals were examined. Southern blot, PCR and sequence analysis reveals that the telomere in chimpanzees and gorillas was not in the same position as the human Xp:Yp telomere. In gorilla two short interstitial blocks of telomere-like repeats were identified, one starting in a position orthologous to the -128 position of the human Xp:Yp telomere. PCR analysis using a primer for the chimpanzee and gorilla subterminal satellite revealed that a block of this subterminal satellite may be present in chimpanzees in a position orthologous to the human Xp:Yp telomere. The human Xp:Yp telomere may be relatively young in evolutionary terms and its location is unique to the human lineage. The terminal regions of the primate X and Y chromosomes have undergone substantial changes over a short period of evolutionary time.

Introduction

Single copy human pseudoautosomal sequences are highly conserved in chimpanzees and segregation analysis reveals that these sequences do not show a sex linked mode of inheritance (Weber *et al.*, 1987, 1988). Thus a pseudoautosomal region has been defined in chimpanzees (*Pan troglodytes*). The probe 29C1 defines the human pseudoautosomal locus DXYS14, 13.4kb from the start of the Xp:Yp telomere and *in situ* hybridisation with the 29C1 probe demonstrated a conserved location at the distal end of the chimpanzee pseudoautosomal region, at Xp22-Xpter and Yq12.2-Yqter (Weber *et al.*, 1988).

Estimates of the divergence between human and chimpanzee non-transcribed sequences are in the order of 1.6%, while estimates of the divergence between human and gorilla non-transcribed sequences are about 1.9% (Miyamoto *et al.*, 1988; Perrin-Pecontal *et al.*, 1992; Mohammad-Ali *et al.*, 1995). Haplotypes identified in the DNA flanking the human Xp:Yp telomere are highly diverged (upto 2.2%, see chapter 3), which may mean that these sequences in humans are evolving at a high rate. If these sequences were also evolving at a similar rate in chimpanzees and gorillas one would anticipate that the divergence of these sequences from humans would be greater than that observed at other loci. If this were the case, then the factors that have influenced the evolution of these sequences in humans could also be operating in chimpanzees and gorillas.

Southern blot analysis carried out by Dr. Nicola Royle, using a probe adjacent to the Xp:Yp telomere containing the proximal three repeat units of the minisatellite, detected smears of hybridising fragments in humans. This observation was consistent with the detection of the terminal restriction fragment of the short arm of the X and Y chromosomes. However, in chimpanzee and gorilla DNAs this probe revealed discrete fragments. In three chimpanzee DNAs a variable number of bands was detected in *Sau3AI* digests. In one DNA the probe revealed three discrete bands of 750bp, 1kb and 4.6kb, in another DNA just the 750bp and 4.6kb fragments were detected and in the third DNA just the 750bp fragment was detected. In a *Sau3AI* digestion of three gorilla DNAs a single 550bp band was detected upon hybridisation to the minisatellite probe. The restriction patterns from chimpanzees and gorillas suggested that the telomere might not be in the same location as in humans. No signal was detected upon hybridisation of *Sau3AI* digested orang-utan (*Pongo pygmaeus*) DNA to the human minisatellite probe.

The subterminal sequences isolated from many organisms are often species specific and show high levels of within species variation in the distribution

between chromosome ends (Brown *et al.*, 1990; de Lange *et al.*, 1990; Louis and Haber, 1990; Ijdo *et al.*, 1992). This variation confirms that the terminal regions of chromosomes are dynamic structures and have a high rate of turnover in evolutionary terms. This is particularly striking when the terminal sequences of human chromosomes are compared with closely related species such as chimpanzees, gorillas and orang-utans. The subterminal repeat sequences of human chromosomes cross hybridise to the genomes of chimpanzees, gorillas and orang-utans (Royle *et al.*, 1994; Hoglund *et al.*, 1995). However, the sequences isolated from adjacent to many chimpanzee telomeres showed no homology to the human subterminal repeat sequences, instead a 32bp subterminal satellite sequence has been isolated (Royle *et al.*, 1994). This was shown by *in situ* and Southern hybridisation to be present in large arrays, at the ends of half of the chimpanzee chromosome arms and virtually all gorilla chromosome arms, and was thought to account for the additional G-bands observed at the ends of gorilla and chimpanzee chromosomes (Yunis and Prakash, 1982; Royle *et al.*, 1994). This subterminal satellite is absent from the human and orang-utan genomes. Additional terminal G-bands have been observed on the short arm of the X and Y chromosomes in gorilla but not in chimpanzees (Yunis and Prakash, 1982) however, fluorescent *in situ* hybridisation revealed that the terminal regions of the short arms of the X and Y chromosomes hybridised to the subterminal satellite in both chimpanzees and gorillas (S. Di and N. J. Royle pers. comm.).

The existence of $(TTAGGG)_n$ at the telomeres of chimpanzee and gorilla chromosomes has been confirmed (Luke and Verma, 1993)(N. J. Royle pers. comm.). But, the physical relationship between the DXYS14 like Xp:Yp pseudoautosomal sequences, the block of subterminal satellite and the telomere repeat array in chimpanzees and gorillas is unknown.

This work. In this chapter I have described the analysis of sequences orthologous to those flanking the human Xp:Yp telomere in chimpanzees, gorillas and orang-utans. The information obtained is discussed from the standpoints of molecular evolution and the evolutionary dynamics of subterminal sequences.

Results

1. Amplification from non-human primate DNAs of sequences orthologous to those between 2 and 1.5kb proximal to the human Xp:Yp telomere

To determine whether sequences between 2kb and 1.5kb from the human Xp:Yp telomere are present in some non-human primate DNAs, the primers TSK8K and TSK8J were used to amplify sequences from chimpanzees, gorillas and orang-utans. The two orang-utan DNAs failed to amplify. Two gorilla DNAs did amplify producing a band of the same size as that obtained from human DNA (Fig. 6.1.). Three chimpanzee DNAs also amplified, the PCR product was approximately 45bp longer than that obtained from human DNAs (Fig. 6.1.). The PCR products were digested with *NlaIV* which detects the -1888 polymorphism in human DNA but this enzyme did not cut the amplified DNA from chimpanzees or gorillas (data not shown). The PCR analysis described here was carried out using the same conditions used for amplifying these sequences from humans, with an annealing temperature of 66°C and discrete PCR products were obtained from both chimpanzee and gorilla DNAs. This suggests that the equivalent sequence is present in chimpanzees and gorillas but this can only be confirmed by sequence analysis.

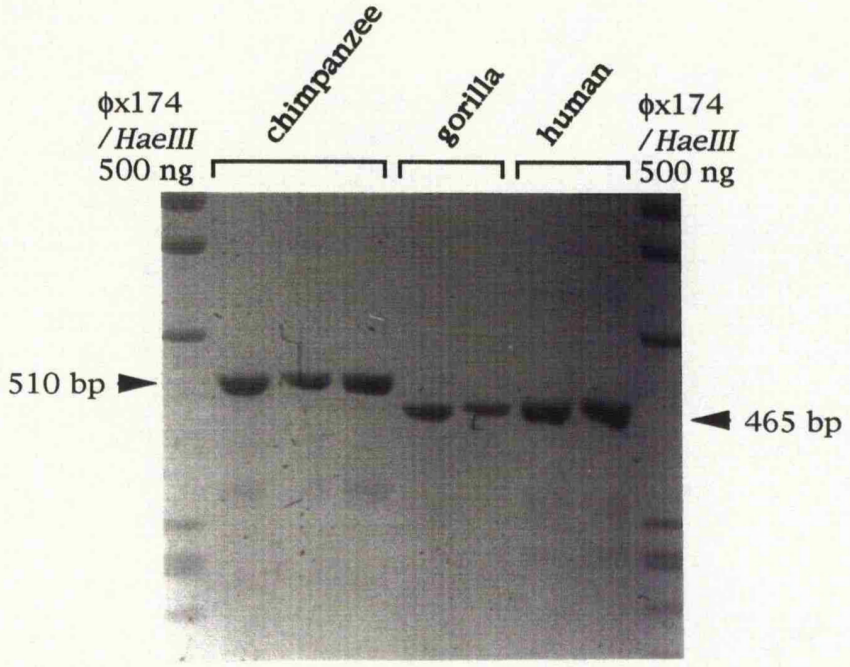
2. Analysis of chimpanzee and gorilla sequences orthologous to the minisatellite flanking the human Xp:Yp telomere

PCR analysis. PCR products obtained from the amplification from chimpanzee and gorilla DNA with primers TSK8C and TSK8F (Fig. 6.2.) were sequenced. Orang-utan DNA does not amplify with TSK8C and TSK8F. The primer TSK8H was designed after comparison of human, chimpanzee and gorilla sequences so that it would anneal to chimpanzee, gorilla and human DNA. The TSK8H primer is oriented 5′ to 3′ towards the telomere 770bp from the human telomere on the proximal side of the minisatellite (Fig. 6.2.). PCR analysis using primers TSK8H and TSK8G revealed in three gorilla DNAs a product smaller than that obtained in humans and chimpanzees. In gorillas this product was approximately 60bp smaller which was accounted for by a lack of one of the minisatellite repeat units, whereas four repeats were present in humans and chimpanzees. Two primers PrimA and PrimB were designed from chimpanzee and gorilla sequence, these primers anneal to the proximal and distal most ends of the minisatellite respectively, both oriented 5′ to 3′ towards the telomere (Fig. 6.2.). Amplification with primers PrimA and TSK8G from a orang-utan DNA revealed a product which was 130bp shorter than that obtained from human and chimpanzee DNAs using the same

Figure 6.1. The amplification of primate DNAs with primers TSK8J and TSK8K.

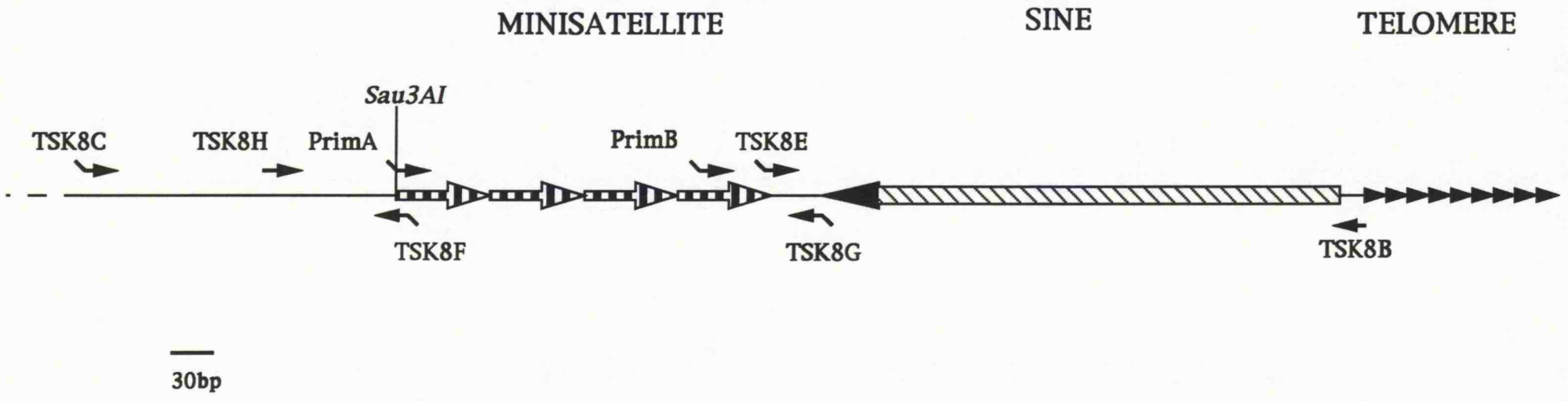
An ethidium bromide stained gel showing PCR products generated from the amplification of chimpanzee, gorilla and human DNA using primers TSK8J and TSK8K.

PCR reactions were carried out using a Geneamp 9600^{TM} thermal cycler (Perkin Elmer). 7µl reactions contained; the standard PCR buffer (see Chapter 2), 1µM of each primer, approximately 100ng of genomic DNA and 0.1 U/µl of Taq polymerase. The reactions were cycled 32 times, each cycle consisted of denaturation at 96°C for 20 seconds, primer annealing at 66°C for 30 seconds and extension at 70°C for 60 seconds. The products of the PCR reaction were resolved by electrophoresis through a 1.5% agarose gel (FMC) in 0.5×TBE.



Chapter 6. Figure 1

Fig. 6.2. Primers used in the analysis of sequences orthologous to the human Xp:Yp pseudoautosomal telomere in non-human primates



primers, this suggests that the minisatellite is also present in orang-utans but contains only two minisatellite repeat units (data not shown).

Sequence analysis. Sequence analysis was carried out on PCR products obtained from the amplification of two chimpanzee DNAs and three gorilla DNAs using primers TSK8H and TSK8G (Fig. 6.2.). A total of 288bp and 251bp were obtained from chimpanzees and gorillas respectively using primer TSK8G (Fig. 6.3.). The sequences from all the chimpanzee DNAs were identical and both the gorilla sequences were also identical. This analysis confirmed that the gorilla DNA contained just three repeat units in the minisatellite. Since this minisatellite shows sequence heterogeneity between the repeat units, it was possible to determine that the third most distal repeat unit was missing in the gorilla DNA (Fig. 6.3.). Both chimpanzees and gorillas showed homology with the Caucasian haplotype A at positions -415, -427, -540, -554 and -652, but at the -544 position in both had a C instead of a G (Fig. 6.3.), however, both the sequences were more diverged from the Caucasian haplotype A than the maximum divergence between the different human haplotypes. Curiously the chimpanzee sequence was more diverged from human (3.5%) than the gorilla sequence (2.8%)(Fig. 6.3.), but given that a relatively small amount of DNA was analysed this is probably not significant. The divergence between the gorilla and chimpanzee sequences was 4.5%. A large proportion of the differences between the human sequence and that of the chimpanzees and gorillas are situated around the -544 and -540 polymorphisms (Fig. 6.3.). These two polymorphisms did not show complete linkage disequilibrium in the West-African population and therefore contributed to the additional heterogeneity in the haplotypes identified in this population.

The sequence analysis confirmed the presence of a *Sau3AI* restriction enzyme site proximal to the minisatellite in chimpanzees and gorillas (Fig. 6.2.). This meant that the single *Sau3AI* band detected in gorillas and the multiple bands in some chimpanzees would extend into the region orthologous to the human Xp:Yp telomere. The presence of multiple bands in some chimpanzees suggests that the *Sau3AI* sites flanking the probe region in chimpanzees are both polymorphic or that this sequence may be duplicated. It also indicates that the telomere in gorillas and chimpanzees is not in the same location relative to the minisatellite sequence as it is in humans.

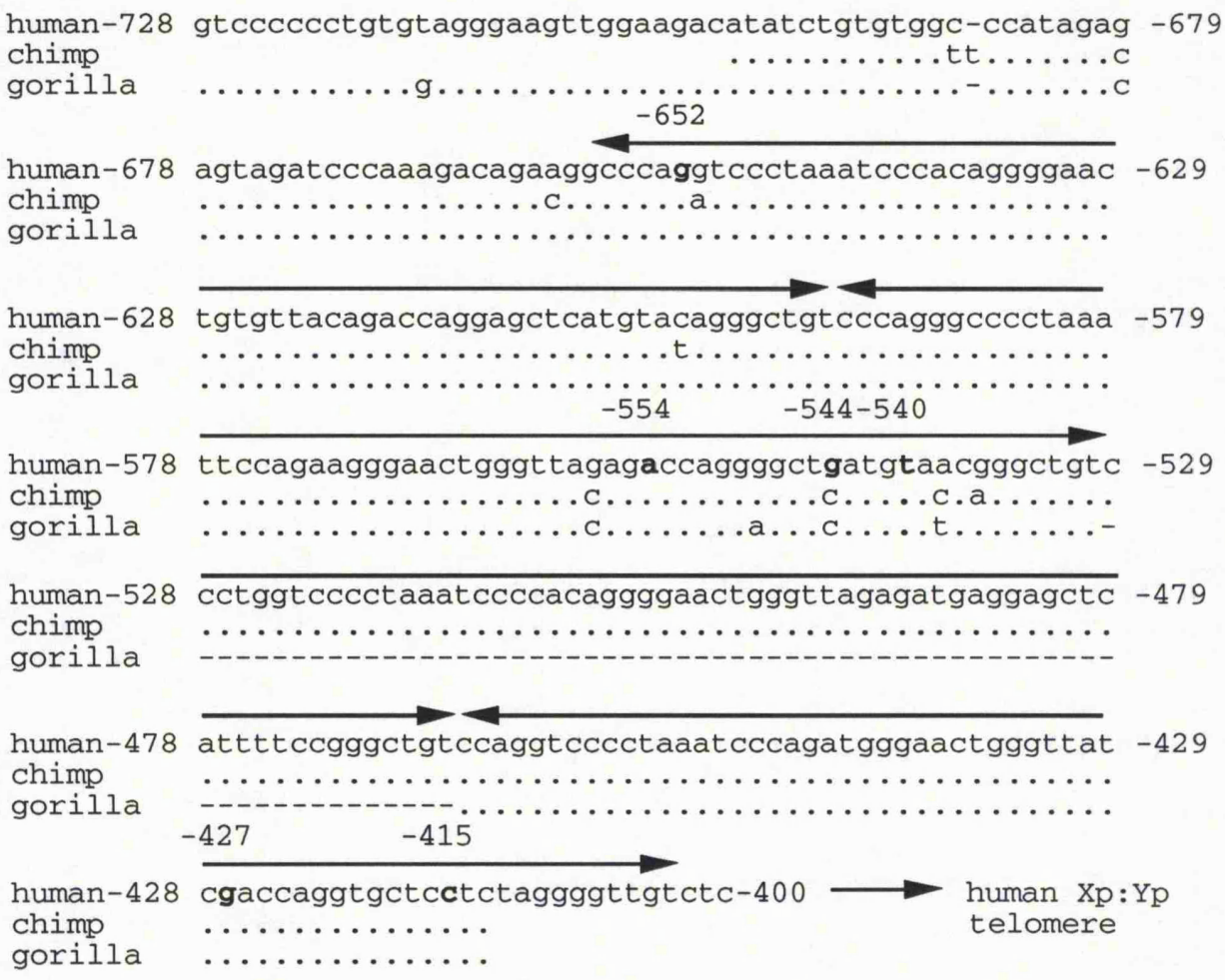
Fig. 6.3. Comparison of chimpanzee and gorilla sequences across the minisatellite flanking the Xp:Yp telomere with the sequence of Caucasian haplotype A.

The sequence on the top line is of the human Caucasian haplotype A, the second and third lines are chimpanzee and gorilla sequences respectively. A position with a '.' represents sequence identity, a '-' represents a missing base and the arrow above the sequence represent the position of the minisatellite repeat units. The bases shown in **bold** are the positions (numbered above) that are polymorphic in humans. The numbering system is based upon the number of bases from the start of the human Xp:Yp telomere.

The chimpanzee sequence shows a total of ten (including one insertion) differences with the human sequence over 288bp this represents a sequence divergence of 3.5%.

The gorilla sequence shows a total of seven (including one deleation) differences with the human sequence over 251bp this represents a sequence divergence of 2.8%. The gorilla sequence is missing the third minisatellite repeat unit.

Comparison of the chimpanzee and gorilla sequences shows a total of ten (including 2 insertions/deleations) differences over 223bp this represents a sequence divergence of 4.5%.



Chapter 6. Figure 3

3. An attempt to detect telomeric sequences in chimpanzee, gorilla and orang-utan DNAs

The initial Southern blot analysis coupled with the sequence analysis of the minisatellite region, suggested that the telomere in chimpanzees and gorillas may not be located at the same position as the Xp:Yp telomere in humans. In order to investigate this further, attempts were made to amplify the telomere repeat array in DNAs from two chimpanzees, two gorillas and one orang-utan. This was carried out by amplification using primers TSK8H with the telomere repeat primers TelH, TelG and TelK which detect TTAGGG, TGAGGG and TCAGGG repeats respectively (see chapter 4). The products were resolved by electrophoresis through a 1% agarose gel, and detected by hybridisation of the Southern blot with a probe consisting of the human minisatellite region (generated by PCR with primers TSK8C and TSK8G). None of the chimpanzee and orang-utan amplified products hybridised to the minisatellite probe (Fig. 6.4.). In gorillas, PCR products amplified with the primer TSK8H and the telomere primers TelH and TelG hybridised to the probe. No products hybridised to the probe in the amplification from gorilla DNA with TSK8H and TelK. Two bands were detected in both DNAs with the TelH primer, the smallest band starts at a size that would be approximately 130bp proximal to the start of the human telomere repeat array (Fig. 6.4.). The smallest band had the appearance of being composed of several fragments and it may be composed of approximately ten telomere-like repeat units. No smear of products such as that observed in the human control was detected in the amplified products from gorilla DNAs (Fig. 6.4.). This suggested, that the bands observed by the amplification with TSK8H and TelH from gorilla DNA were due to small interstitial blocks of telomere-like repeats. Again it seemed most likely that the Xp:Yp telomere in chimpanzee and gorilla DNAs was not in the same position relative to the flanking DNA found in humans. Based on the assumption that the PCR reactions would not detect products greater than 5kb, the Southern blot shown in figure 6.4. suggests that the telomere in chimpanzees and gorillas must be greater than 4kb distal to the start of the Xp:Yp telomere in humans.

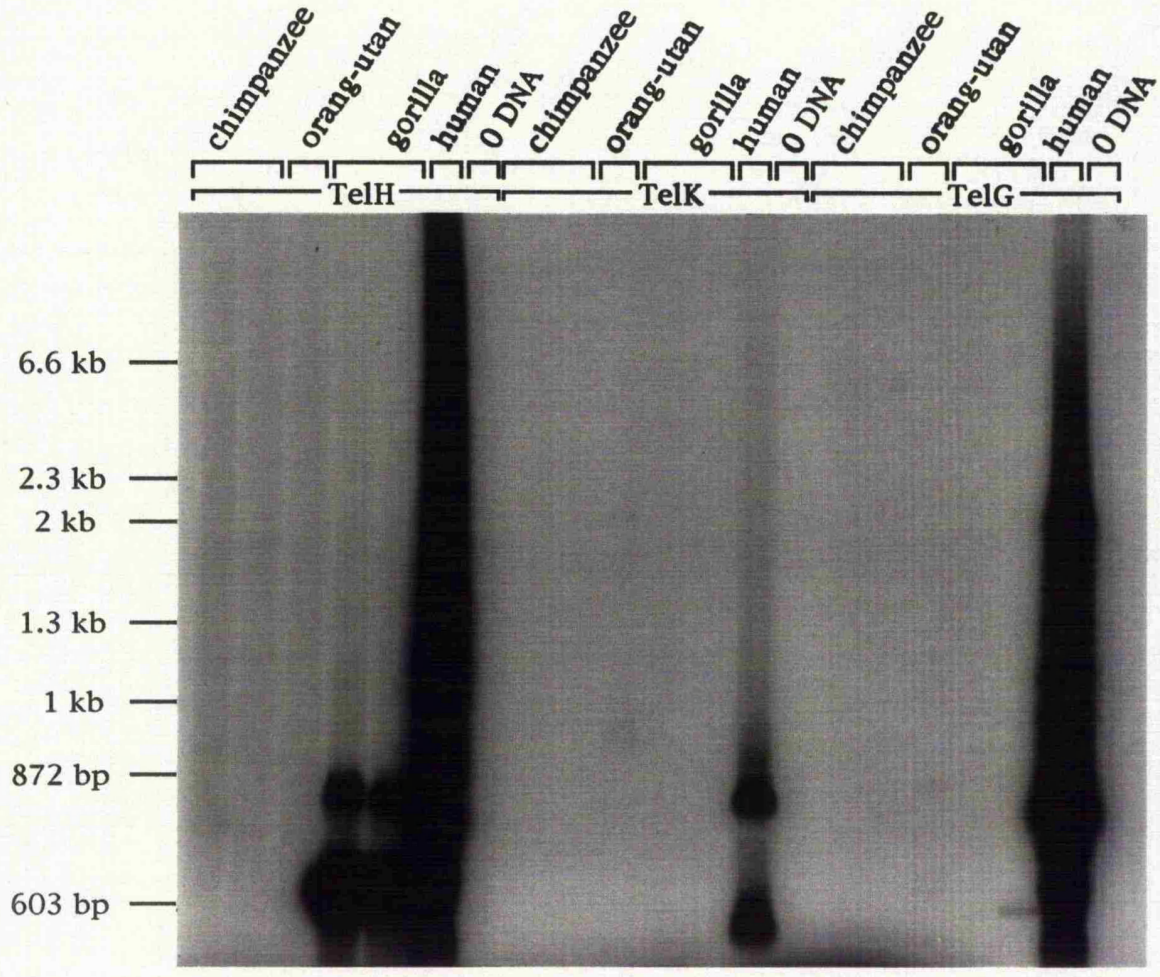
4. Analysis of gorilla sequences orthologous to the SINE flanking the human Xp:Yp telomere

The bands observed after amplification with TSK8H and TelH from gorilla DNA were amplified to a visible level on an ethidium bromide stained agarose gel. This produced a smear with a superimposed band; the band was purified and sequenced using the primer TSK8E. It was possible to sequence this PCR product completely. The sequence obtained demonstrated in gorillas the existence of sequences with

Fig. 6.4. The amplification of primate DNAs with primers TSK8H and three telomere or variant repeat primers.

A Southern blot of PCR amplified products from chimpanzee, orang-utan, gorilla and human genomic DNAs using primers TSK8H and the telomere repeat primers shown was hybridised to the human minisatellite probe. Only products from human and gorilla DNA hybridised to the probe. Two strongly hybridising bands were observed upon amplification with TelH with gorilla DNA and one faintly hybridising band was observed after amplification with TelG from one of the gorilla DNAs.

The PCRs were carried out using a Geneamp 9600^{TM} thermal cycler (Perkin Elmer). The 10µl reactions contained 1µM of TSK8H with an equal concentration of one of the telomere repeat primers, the standard PCR buffer and 0.1 U/µl of Taq polymerase. The reactions were cycled 22 times, each cycle consisted of denaturation at 96°C for 50 seconds, primer annealing at 66°C for 30 seconds and extension at 70°C for 10 minutes. Products were resolved in 1% HGT agarose (FMC) gel and the Southern blot was hybridised to the minisatellite probe in the 'blotto' hybridisation mix and the filter washed to a stringency of 0.2×SSC/0.1%SDS.



TelH =TTAGGG

TelK =TCAGGG

TelG =TGAGGG

Chapter 6. Figure 4

homology to the SINE flanking the human Xp:Yp telomere. The distal end of the gorilla PCR product terminated with the sequence GCAGGGTTAGGG, which defines the point of divergence from the SINE sequence and the start of the interstitial array of TTAGGG and variant repeats in the gorilla (Fig. 6.5.). The human sequence at this position bears no resemblance to TTAGGG and nor do published sequences from the family of SINEs (La Mantia *et al.*, 1989). The start of this interstitial block of repeats corresponds to a position orthologous to the -128 position at the human Xp:Yp telomere. The gorilla sequence shows a 6.0% divergence from Caucasian haplotype A (Fig. 6.5.). The haplotype obtained from gorilla DNA is very similar to the Caucasian haplotype A at the polymorphic positions that define this haplotype in humans, except at the -338 and -333 positions where the gorilla sequence is the same as the West-African haplotype A (Table 6.1.).

When the sequence information from the minisatellite and the SINE from gorillas is combined and compared to the sequence from the Caucasian haplotype A the divergence is 4.3%.

5. Sequences orthologous to the human Xp:Yp telomere

Recently a subterminal satellite has been isolated which is present in large blocks adjacent to the telomeres in gorillas and chimpanzees, but absent from human and orang-utan genomes (Royle *et al.*, 1994). Fluorescent *in situ* hybridisation (FISH)(Royle *et al.*, 1994) demonstrated the presence of the subterminal satellite at the distal end of Xp:Yp in chimpanzees and gorillas (S. Di and N. J. Royle pers. comm.). To investigate the physical relationship of the subterminal satellite to the Xp:Yp minisatellite in chimpanzees, gorilla and orang-utans, a PCR primer was designed to anneal to the subterminal satellite but oriented 5′ to 3′ towards the centromere (Cht7B). Amplification with Cht7B and PrimA (located at the proximal end of the minisatellite, see figure 6.2.) followed by Southern blot hybridisation of the PCR products to a probe specific for the minisatellite, revealed a smear of hybridising fragments in chimpanzee amplified DNA (Fig. 6.6.). The amplified products had a periodicity which approximated to 32bp, which is the subterminal satellite repeat unit length (Royle *et al.*, 1994). The shortest fragments of the smear were 600bp, which places the start site of the subterminal satellite at a position orthologous to the -70 position of the human Xp:Yp telomere. No hybridising products were detected in gorilla and human DNA, amplification from orang-utan DNA revealed two faintly hybridising fragments after 30 cycles of PCR, these are presumably due to some non-specific amplification. Amplification with Cht7B and PrimB (located at the distal end of the minisatellite, see figure 6.2.) from chimpanzee DNA produced a

Fig. 6.5. Comparison of gorilla sequences across the SINE immediately adjacent to the human Xp:Yp telomere with the sequence of Caucasian haplotype A.

```
                -338 -333                                    -298-297
human-345 tgaagacgcggaacctcgcggtgagtgttacagttcttaaaggtggcatg-296
gorilla   ....t..a....c..............c.....................

human-295 tccggagtttgtttcttctgatgttcagatgtgttctgagtttcttcttt-246
gorilla   ..........-a.c.....................g..........c..

human-245 ctggtggggttgtggtctcactggctcaggagtgaagctgcagacctttg-196
gorilla   ................................................

                          -176                          -147-146
human-195 cggtgagtgtcacagctcagaaaggcagtgtggacccaaagagtgagcag-146
gorilla   .a.................................a..........c...

        -145
human-145 tagcaagatttattgcaaagagtgaa-120 ──► the human Xp:Yp
gorilla   ..-.............gggttaggg          telomere
```

There is a total of thirteen (including two deletions) differences in 217bp which represents a sequence divergence of 6.0%.

The '.' below the human sequence represent sequence identity in gorillas and '-' represents missing bases. The numbering system is based on the distance from the start of the human Xp:Yp telomere. The sequences shown in **bold** are the polymorphic positions identified in humans, the human sequence shown is that of the Caucasian haplotype A.

TABLE 6.1. Comparason of Caucasian and West-African haplotype A with the sequences obtained from gorilla and chimpanzee DNAs

																			→ Telomere
position from first telomere repeat in humans	-842 *DdeI* +/-	-652 *Ava II* +/-	-554	-544	-540	-427 *TaqI* +/-	-415 *Mbo II* +/-	Δ 373-363[a]	-338	-333	-298	-297	-176 *DdeI* +/-	-147	-146	-145	-73	-30 ARMS	-13 ARMS
Haplotypes from sequence analysis.																			
Caucasian Haplotype A	C	G	A	G	T	G	C	+	G	A	A	T	G	A	G	T	C	T	T
chimpanzee haplotype	nd	G	A	C	T	G	C	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
gorilla haplotype	nd	G	A	C	T	G	C	nd	A	C	A	T	G	A	G	T	nd	nd	nd
West-African Haplotype A	C	G	A	C G	C T	G	C	-	A	C	A	T	G	G	G	C	G	T A	nd

[a] The 10 bases between -363 and -373 are present (+) or absent (-) as indicated.

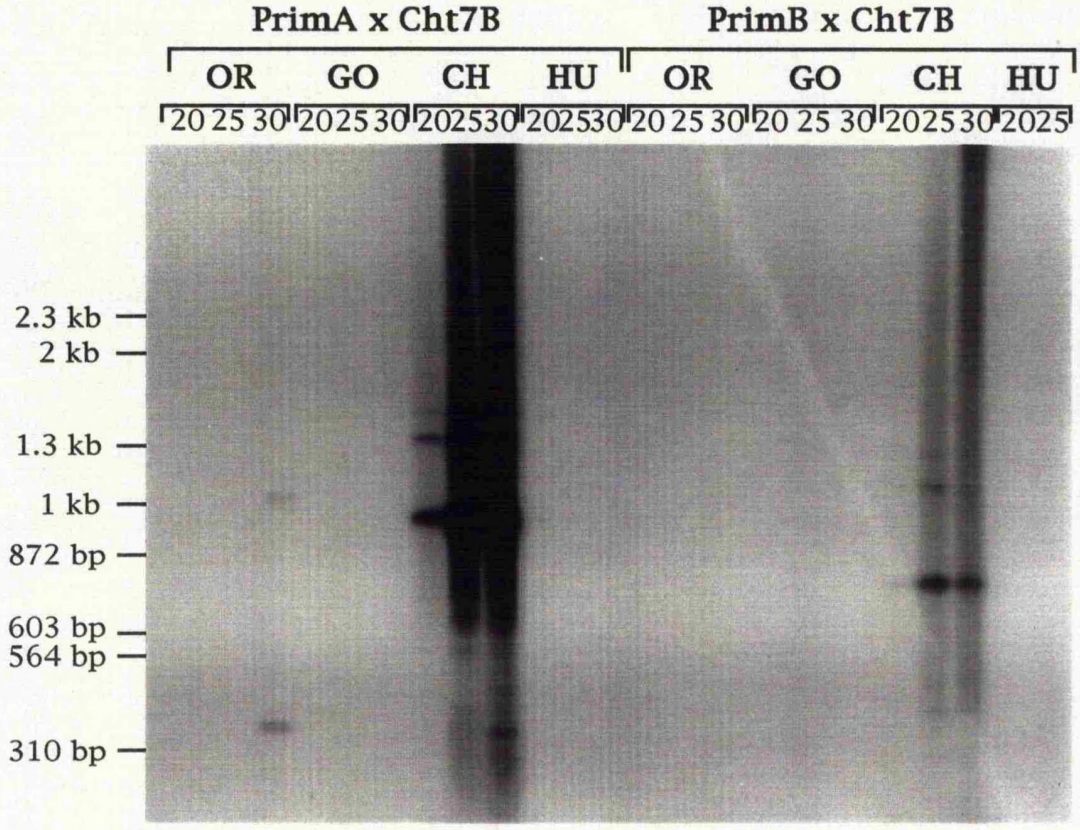
nd=sequence not determined.

Fig. 6.6. PCR amplification of primate DNAs with primers specific to the Xp:Yp minisatellite and the 32bp subterminal satellite.

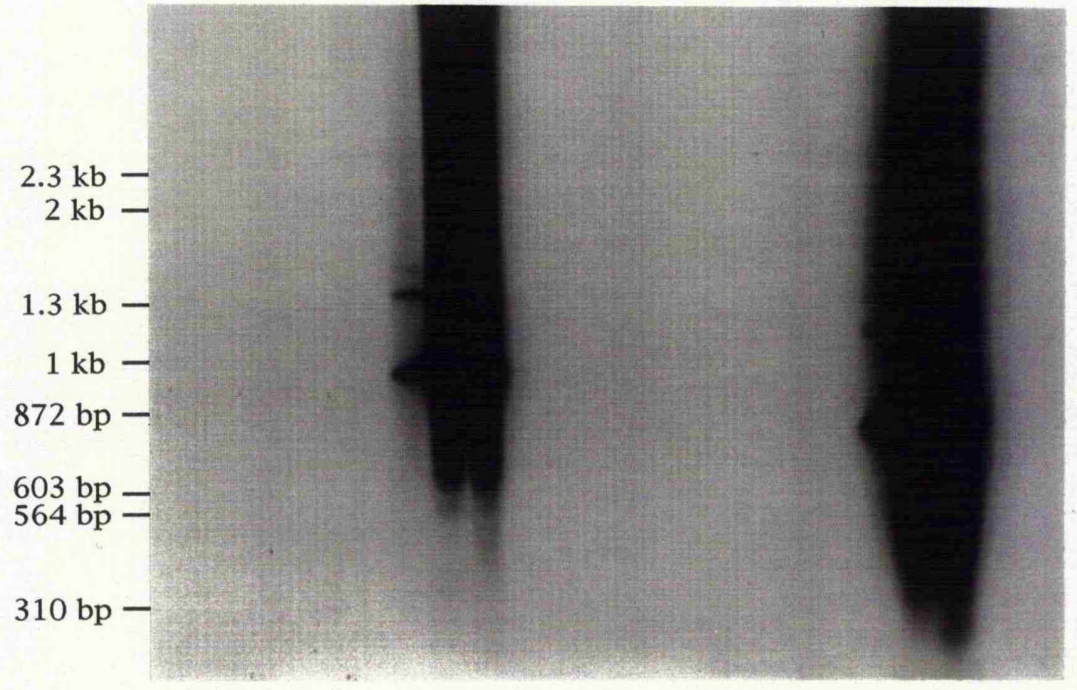
A Southern blot of PCR products from orang-utan (OR), gorilla (GO), chimpanzee (CH) and human (HU) genomic DNAs after amplification with primers specific to the Xp:Yp minisatellite (PrimA or PrimB) and a primer for the chimpanzee and gorilla subterminal satellite. The Southern blot was hybridised to a DNA probe for the human Xp:Yp minisatellite (Top) or to a DNA probe for the SINE adjacent to the human Xp:Yp telomere (Bottom).

The 7µl PCRs were carried out using 1µM of Cht7B used in conjunction with an equal concentration of PrimA or PrimB, the standard PCR buffer (see Materials and Methods) with 0.1 U/µl of Taq polymerase. The reactions were cycled using a Geneamp 9600^{TM} thermal cycler (Perkin Elmer), each cycle consisted of denaturation at 96°C for 20 seconds, primer annealing at 66°C for 30 seconds and extension at 70°C for 60 seconds. The number of cycles used is shown above each lane. Hybridisation was carried using the 'blotto' hybridisation mix and the filter washed to a stringency of 0.1×SSC/0.1%SDS.

Blot hybridised to the minisatellite probe



Blot hybridised to the SINE probe

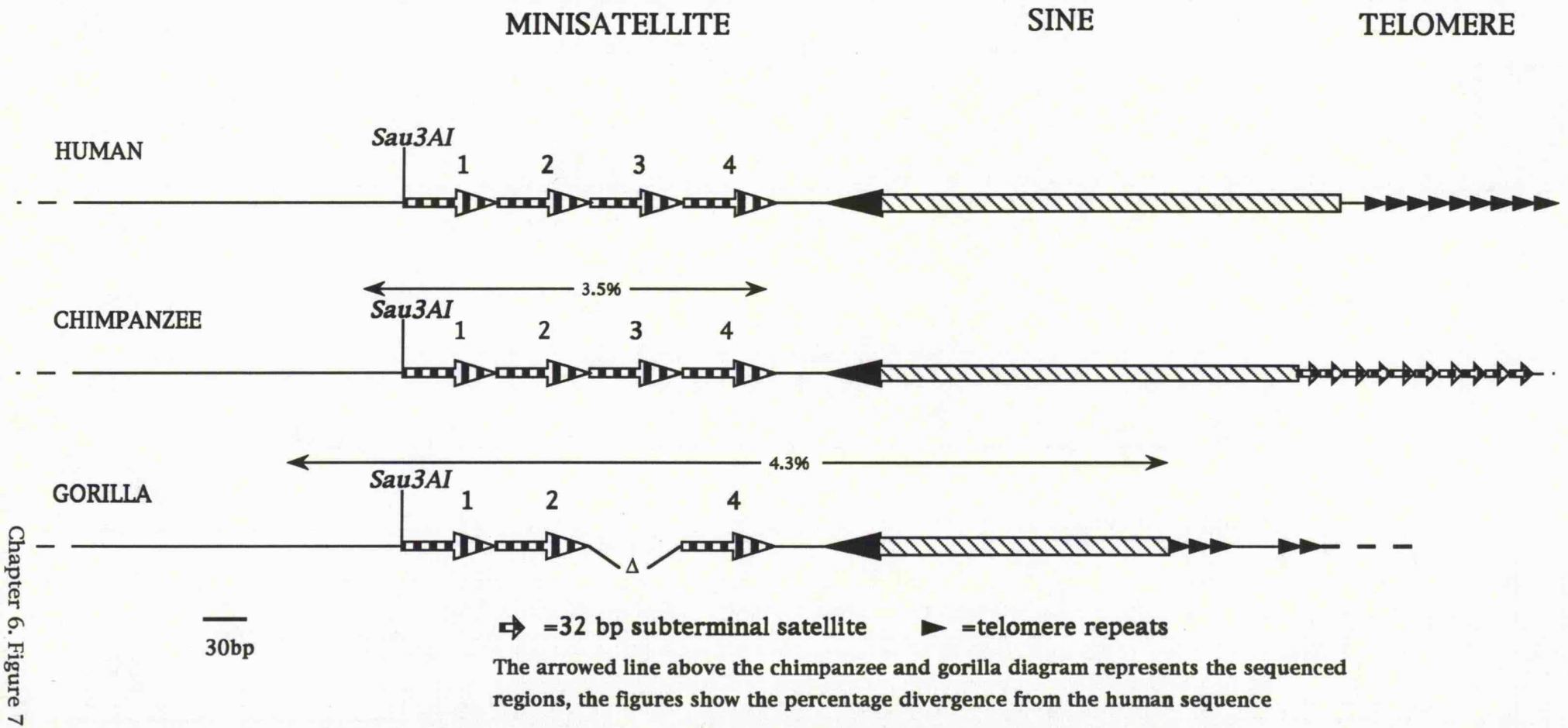


Chapter 6. Figure 6

similar set of fragments, as expected shortened by 230bp. The minisatellite probe hybridised to these fragments with less intensity because these products had only 47bp of homology to it (Fig. 6.6.). Hybridisation with a probe specific to the SINE produced equally intense hybridisation with the products generated using either primers (PrimA and PrimB) (Fig. 6.6.). This analysis, demonstrated the presence of the SINE distal to the Xp:Yp minisatellite in chimpanzees and showed that a block of subterminal satellite is present at the same location as the telomere in humans.

The analysis presented in this chapter reveals that sequences similar to those flanking the human Xp:Yp telomere are present in the genomes of chimpanzees, gorillas and orang-utans. These sequences have a greater divergence from the human sequence than that detected at other loci, such as the β-globin locus (Perrin-Pecontal *et al.*, 1992). The sequence between 2.0kb and 1.5kb from the human Xp:Yp telomere appears to be present in chimpanzees and gorillas, the chimpanzee sequence is 45bp longer than that derived from humans and chimpanzees. The minisatellite is present in all three species analysed, in chimpanzees, as in humans, four repeat units of the minisatellite are present, in gorilla only three repeat units are present (Fig. 6.7.) and in orang-utan there are probably only two repeat units. In humans there is a truncated version of a SINE immediately adjacent to the Xp:Yp telomere, in chimpanzees this appears to be truncated even further by the presence of an array subterminal satellite, the exact start of the subterminal satellite requires definition by sequence analysis (Fig. 6.7.). In gorillas, two small interstitial blocks of telomere-like repeat units are present, one starting in a position that corresponds to the -128 position of the human telomere. Currently it is unknown what sequences occur distal to the interstitial array of TTAGGG and variant repeats in the gorilla genome.

Fig. 6.7. Showing the structures present in chimpanzees and gorillas orthologous to those present at the human Xp:Yp pseudoautosomal telomere junction region



The arrowed line above the chimpanzee and gorilla diagram represents the sequenced regions, the figures show the percentage divergence from the human sequence

Discussion

The sequence information obtained from chimpanzee and gorilla DNA show homology to the human Caucasian haplotype A (Table 6.1.) at the positions identified as polymorphic that define haplotype A (chapter 3). The genome of the ancestral species which gave rise to the human, chimpanzee and gorilla lineages must have contained haplotype A. There was no evidence of polymorphism from the sequence analysis in chimpanzees and gorillas, but this analysis was only carried out in a few individuals. In addition, chimpanzee and gorilla samples came from the captive animals and may have been inbred, resulting in a lower level of sequence variation.

One possibility to explain the highly diverged haplotypes adjacent to the human Xp:Yp telomere, is that it is caused by a conversion from one locus fixed for one haplotype to the locus adjacent to the Xp:Yp telomere fixed for a different haplotype (see chapter 3). There is no evidence of a duplication of this locus in humans. The analysis in this chapter does not provide any evidence of a duplication of sequences orthologous to the DNA flanking the human Xp:Yp telomere in the chimpanzee or the gorilla genomes. However, additional bands were observed during Southern blot analysis on *Sau3AI* digests of chimpanzee DNAs. The length of these fragments means that the subterminal satellite in chimpanzee must contain *Sau3AI* sites. If this is the case, these fragments could arise from polymorphic *Sau3AI* sites in the subterminal satellite. This subterminal satellite contains heterogeneity in the repeat unit sequence and some variant repeat units of this subterminal satellite contain *Sau3AI* restriction enzyme sites (Royle *et al.*, 1994). However, allelic variation in the subterminal satellite cannot explain the presence of the three *Sau3AI* fragments observed in one chimpanzee DNA. Additional analysis of this region will be required to obtain restriction maps from a number of chimpanzee DNAs to determine the extent of allelic variation in these *Sau3AI* RFLPs and whether there is any evidence of a second locus.

Sequence information obtained from chimpanzee and gorilla DNA orthologous to the human sequences adjacent to Xp:Yp telomere, allowed the divergence from the human sequences to be determined. The chimpanzee sequences are diverged from the human Caucasian haplotype A across the minisatellite by 3.5%. In gorilla, sequence was obtained across the minisatellite and SINE and the divergence across this region from the human Caucasian haplotype A is 4.3%. Over this region the sequence divergence is higher between African apes than has been observed at other loci; for example, divergence between human and chimpanzee sequences at the β-globin locus is 1.62% and between humans and

gorillas is 1.85% (Perrin-Pecontal *et al.*, 1992). The sequence divergence of 4.5% between the chimpanzee and gorilla sequences orthologous to the DNA flanking the human Xp:Yp telomere, is again higher than that observed at other loci (Miyamoto *et al.*, 1988; Mohammad-Ali *et al.*, 1995). The high sequence divergence between the haplotypes in the DNA flanking the human Xp:Yp telomere suggests that this region may be evolving rapidly. The higher sequence divergence observed in the sequences from chimpanzees and gorillas orthologous to those flanking the human Xp:Yp telomere, suggests that these loci in chimpanzees and gorilla are also evolving at a higher rate than the rest of the genome. It was suggested in chapter 3 that the high frequency of sequence polymorphism observed in humans may be a consequence of its location immediately adjacent to a telomere, perhaps preventing access to these sequences by proteins involved in repair mechanisms. However, the analysis in this chapter demonstrates that these sequences in chimpanzees and gorillas are not immediately adjacent to a telomere, suggesting that the rate of evolution of this locus is not necessarily a consequence of its location immediately adjacent to a telomere, but may be due to other factors such as selection for polymorphism in a gene at or near this locus (see chapter 3).

PCR and sequence analysis of the gorilla sequences orthologous to the human SINE flanking the Xp:Yp telomere revealed that the telomere array was not in the same position. Instead two small interstitial blocks of telomere-like repeats were present, the most proximal block was at a position equivalent to the -128 position of the human Xp:Yp telomere (Fig. 6.7.). It is not known what sequences are present distal to these interstitial arrays of telomere-like repeat units. However, fluorescent *in situ* hybridisation (FISH) with a probe for the subterminal satellite revealed that the ends of the short arms of the X and Y chromosomes of gorilla contained the chimpanzee and gorilla specific subterminal satellite. This was not detected by PCR analysis distal to the minisatellite, this could be because the subterminal satellite was too far from the minisatellite for amplification using PCR, or because the subterminal repeat unit sequence is sufficiently diverged to prevent amplification with the repeat unit primer used. In the chimpanzee, a block of subterminal satellite DNA with 32bp periodicity was present at a position orthologous to the -70 position of the human Xp:Yp telomere (Fig. 6.7.). The exact start site of the array of subterminal satellite requires definition by sequence analysis. This is because the start site of -70 was defined by an error prone estimate of the size of the smallest fragment in a smear of fragments generated by amplification with the subterminal satellite specific primer. The subterminal satellite has been detected at two interstitial sites and at the end of half of the chimpanzee chromosome arms and at the ends of most gorilla chromosome arms (Royle *et al.*, 1994). It is likely that the genome of the ancestral species that gave

rise to the human, chimpanzee and gorilla lineages contained this subterminal satellite. This satellite was then lost from the human lineage and expanded in the chimpanzee and gorilla lineages (Royle *et al.*, 1994). It is possible that the genome of the ancestral species contained the satellite in the subterminal regions of the pseudoautosomal region, which was then lost in the human lineage and the Xp:Yp telomere was formed. This would mean that this telomere in humans is relatively young in evolutionary terms and its location specific to the human lineage. It is also possible that the ancestral species did not have this satellite at this chromosome end, and the X and Y chromosomes in the chimpanzee genome acquired it by exchange events between non-homologous chromosomes. However, this possibility may be less likely given that the human Xp:Yp pseudoautosomal region is unique in the human genome and conserved in the chimpanzee genome, and so may not have been involved in exchange events between non-homologous chromosomes which gave rise to the subterminal sequences shared between most human chromosome ends.

Conclusions

The analysis presented in this chapter demonstrates that sequences similar to the DNA flanking the human Xp:Yp telomere exist in the chimpanzee and gorilla genomes. However, these sequences are highly diverged, more so than that detected in other regions of the genome. The sequences detected in chimpanzees and gorillas are similar to the human haplotype A at the polymorphic positions that define haplotype A. This means that the genome of ancestral species that gave rise to the human, chimpanzee and gorilla lineages may have contained a haplotype A like sequence. This analysis demonstrates that the telomere in chimpanzees and gorillas is not in the same position relative to the DNA flanking the human Xp:Yp telomere. In gorillas there are two small interstitial blocks of telomere repeats. In chimpanzees there is an array of subterminal satellite 32bp repeats present in the region orthologous to the human Xp:Yp telomere. This suggests that the human Xp:Yp telomere may in evolutionary terms be relatively young and its location is unique to the human lineage. The terminal regions of the X and Y chromosomes of primates have undergone substantial changes over a relatively short period of evolutionary time, this highlights the dynamic nature of the terminal regions of linear chromosomes.

Chapter 7

FINAL DISCUSSION AND FUTURE WORK

DISCUSSION

Much of the work presented in this thesis has been discussed in the relevant chapters. In this final discussion I bring together the information and points raised in the other chapters.

In order to investigate the turnover of sequences within the Xp:Yp pseudoautosomal telomere a system for the analysis of sequence variation within a single telomere repeat array was developed. The Xp:Yp telomere was chosen because it was thought to be a unique chromosome end in the human genome which would facilitate the development of a PCR based technique. During the course of the development of this technique some interesting phenomena were revealed in the DNA immediately flanking the telomere.

The main findings of this work were;

1. There is a high frequency of sequence polymorphism in the 1kb of DNA immediately flanking the Xp:Yp telomere, this coupled with strong linkage disequilibrium results in a limited number of highly diverged haplotypes.

2. The divergence between the human Xp:Yp flanking haplotypes is similar to the average divergence between human and chimpanzee sequences in other regions of the genome.

3. The DNA sequence between 1.8kb and 2.0kb from the Xp:Yp telomere shows 87-88% similarity with several uncharacterised cDNA clones.

4. Sequences orthologous to those flanking the human Xp:Yp telomere show a higher level of divergence in chimpanzees and gorillas than that observed between these three species at other loci.

5. The Xp:Yp telomere of chimpanzees and gorillas is not in the same location as the telomere in humans. In chimpanzees, there is an array of subterminal satellite

present at a position orthologous to the human telomere, this subterminal satellite may also be present at this position in gorillas. This implies that the human Xp:Yp telomere is unique to the human lineage and may therefore be relatively young in evolutionary terms.

6. The distribution of TTAGGG, TGAGGG and TCAGGG repeats in the proximal 120 repeats of the human Xp:Yp telomere is highly polymorphic with a high underlying mutation rate.

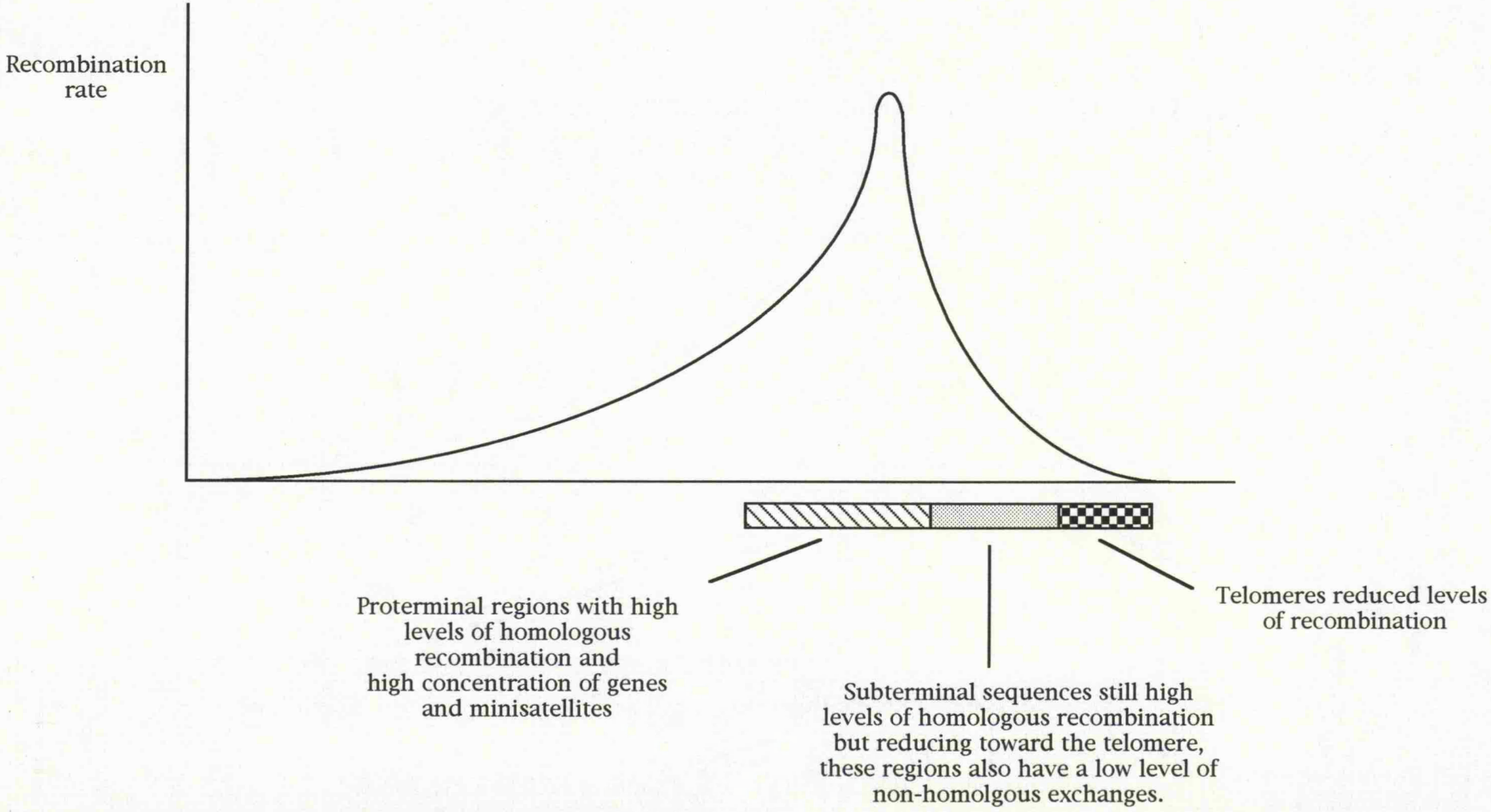
7. The strong linkage disequilibrium observed in the DNA flanking the Xp:Yp telomere extends into the first few repeats of the telomere repeat array. The Xp:Yp telomere appears to evolve along haploid lineages, therefore the predominant mutational mechanism underlying the observed variation must be intra-allelic.

8. The Xp:Yp telomere repeat array starts at a well defined site, but a very rare Caucasian variant exists, in which the telomere starts 3bp distal to the start site of all the other telomeres analysed.

Variation in the DNA flanking the Xp:Yp telomere

Homologue identification during meiotic prophase and the initiation of chromosome pairing is thought to occur towards the ends of human chromosomes (Wallace and Hulten, 1985). In addition, there is an increase in recombination towards the ends of human chromosomes (NIH/CEPH, 1992). The presence of families of subterminal sequences at the ends of most human chromosomes and their polymorphic distribution, suggests that homologue recognition may occur internally to these sequences (Brown *et al.*, 1990; de Lange *et al.*, 1990; Ijdo *et al.*, 1992). The extraordinary length polymorphism detected in the start site of the 16p telomeres suggests that homologue recognition must also occur internally to the point of divergence of the different alleles (Wilkie *et al.*, 1991). It is not known whether the increase in recombination towards the ends of human chromosomes correlates with the presence of subterminal sequence families. Therefore it may be possible that, distal to the point of homologue recognition in the regions of subterminal sequence families, there is a decrease in meiotic recombination such as that observed at 16p (Fig. 7.1.). The long range characterisation of chromosome ends, by including the genetic and physical maps to locate the position of subterminal sequence families relative to genetic markers, would determine whether the regions of chromosomes covered by the subterminal sequence families were proficient in recombination. It would not be

Fig. 7.1. Speculative model of recombination rates along a human chromosome.



expected that the unique chromosome ends in the human genome that do not contain subterminal sequence families would behave in a similar manner to those that do contain these sequence families. However, for the strong linkage disequilibrium and sequence divergence observed at the Xp:Yp telomere to be maintained, a suppression of recombination in this region may be required. Information on recombination rates in terminal regions of the Xp:Yp chromosome arms could be generated by a comprehensive screen for recombination between DXYS14 (the most distal marker on the genetic map) and the telomere.

The sequence divergence of the haplotypes flanking the Xp:Yp telomere suggests that they must be either ancient or that this region has evolved at a faster rate than other regions of the genome. The faster rate of evolution of these sequences could be due to a position effect on this locus because of the proximity of the telomere, for example, telomere binding proteins or an altered chromatin structure could reduce access to this DNA by the mechanisms involved in repair. However, position effects on gene expression have not been observed in humans (Bayne *et al.*, 1994) and the chromatin structure of the proximal regions of human telomeres is similar to the rest of the genome (Tommerup *et al.*, 1994). In addition, the sequences in chimpanzees and gorillas orthologous to those flanking the human Xp:Yp telomere displayed a divergence higher than that observed at other regions of the genome, but these sequences are not adjacent to a telomere in either species. Furthermore, an increased base substitution rate as a result of lower levels of repair cannot account for the presence of strong linkage disequilibrium. To establish the linkage disequilibrium in the presence of a high base substitution rate every time a new sequence variant occurred, the pre-existing haplotype in which it occurred would have to be lost and the new rare variant become common in the population. These processes would have to have occurred multiple times to account for the observed sequence divergence between the haplotypes. Instead, it seems more likely that these sequences are ancient, perhaps as old as the divergence of chimpanzees or even gorillas from the human lineage. The maintenance of this polymorphism over such a long period of time cannot be explained by neutral theory (Nei, 1987) and therefore it seems likely that selection is or has been involved. A balancing selection mechanism similar to that proposed for the maintenance of polymorphism at the HLA loci, could operate to maintain polymorphism at the Xp:Yp telomere. If selection is influencing the evolution of this locus then there must be a gene at or near the telomere. Tantalisingly, sequence similarity of 87-88% was detected to five cDNA sequences from uncharacterised genes in the most proximal 200bp of the 2kb of sequence immediately adjacent to the Xp:Yp telomere. The hypothetical gene would have to be one that displayed phenotypes on which

balancing selection mechanisms could operate, such as those involved in recognition of pathogens.

The divergence of the orthologous chimpanzee and gorilla sequences from those flanking the human Xp:Yp telomere is higher than that observed in other non-conserved regions of the genome. The high divergence could be due to some mechanism actively promoting genetic diversity, again this implies that some kind of selection mechanism may be operating. However, the high sequence divergence could also be a result of a position effect not exerted by the presence of a telomere but due to the proximity to a block of subterminal satellite. In these two species the subterminal satellite could, like other satellite DNA (Jones and Corneo, 1971), be heterochromatic and perhaps reduce the access of DNA repair mechanisms to these sequences, so resulting in an elevated mutation rate.

An alternative hypothesis to explain the presence of these highly diverged haplotypes in the DNA flanking the human Xp:Yp telomere was presented in chapter 3. This involved the conversion from a locus fixed for one haplotype to the Xp:Yp telomere fixed for another haplotype, resulting in two diverged haplotypes adjacent to the Xp:Yp telomere. This appears unlikely, since there is no evidence of a duplicated locus in the genomes of humans, chimpanzees or gorillas. In addition one would expect just two haplotypes to arise from such a conversion event, but more than two haplotypes have been detected. Even if such a conversion event did occur, it would have to have occurred early in the human lineage. Therefore, one would then still have to explain the continued existence of these haplotypes in modern human populations.

Variation in the Xp:Yp telomere

The distribution of different types of telomere repeat within the Xp:Yp telomere is extremely variable, such that heterozygosity estimates are in excess of 99%. The strong linkage disequilibrium observed in the DNA flanking the telomere appears to extend into the first few repeats of the telomere. The allelic structures of different telomeres can be arranged according to similarity to form groups of related alleles, these groups of related alleles share the same haplotype in the flanking DNA. This suggests that these sequences evolve along haploid lineages, therefore, the predominant mechanism involved in the evolution of these sequences must be intra-allelic and the most likely mechanism is thought to be slippage during replication (Levinson and Gutman, 1987)(discussed in chapter 5).

Mechanisms that prevent the spread of variant repeats throughout the telomere may be involved, for example the action of telomerase extending a shortened

telomere or some functional constraint. Functional constraints could include, the reduction in the affinity of telomere binding proteins for telomere sequences, which could affect telomere stability. Telomere binding activity has been detected in humans, some of which is highly specific to the TTAGGG repeat sequence and will not bind some variant repeats (Zhong *et al.*, 1992; Chong et al., 1995). The introduction of variant telomere repeat units into proximal locations within the telomeres of *Tetrahymena* had little effect (Yu and Blackburn, 1991), but if the variant repeats occurred at the terminus of the telomere, phenotypic changes were observed. These changes included, changes in telomere length, impaired cell division and cellular senescence (Yu *et al.*, 1990). Since the distribution of telomere and variant repeats in the proximal 120 repeats of the human Xp:Yp telomere is so variable, functional constraints on their evolution must be minimal and probably only prevent the spread to more distal locations.

Telomere variant repeats are present at many other human telomeres, as observed from sequence analysis (Brown *et al.*, 1990; Royle *et al.*, 1992), hybridisation (Allshire *et al.*, 1989) and TVR-PCR on the telomere of 7q (J. Coleman pers. comm.). These variant repeats presumably arose by mutation and have spread to other telomeres via exchanges between non-homologous telomeres, or by independent mutation events in each telomere. However, it is interesting that the majority of variant repeats only differ in the TTA portion of the TTAGGG telomere repeat and are conserved in the GGG portion, also the 6bp periodicity of the telomere repeat is conserved among the majority of variant repeats. This may reflect functional constraints acting only upon the GGG portion of the telomere repeat, again this could be due to the specificity of telomere binding proteins, or formation of secondary DNA structures. Alternatively, the conservation of the general telomere repeat structure among variants could also be accounted for if the human Xp:Yp telomere is relatively young in evolutionary terms and has not had time to degrade further as a result of mutation in the proximal end of the telomere repeat array. The human Xp:Yp telomere is unique in its location to the human lineage and the orthologous location in chimpanzees and possibly gorillas contains an array of 32bp subterminal satellite repeats. The progenitor to the human, chimpanzee and gorilla lineages probably contained this subterminal repeat sequence (Royle *et al.*, 1994) and it may have been present at the terminal regions of the Xp:Yp chromosome arms (chapter 6). If this is the case, then this satellite was lost in the human lineage and a new telomere seeded at this location. Preliminary work on the isolation of sequences adjacent to telomeres in pigs suggests that the start of the telomere repeat arrays are highly degraded such that it is often difficult to define the start of the telomere (E. Scanlan pers. comm.). This is in contrast to the Xp:Yp telomere and other human telomeres (Brown *et al.*, 1990; Royle *et al.*, 1992)

where the start of the telomere repeat array is well defined. The proximal regions of the telomeres from pigs may be older in evolutionary terms than human telomeres and have degraded further from the consensus sequence. The telomeric alleles devoid of variant repeats, that were detected in all the populations analysed, could be descendants of the original telomere that have not acquired any variant repeat units.

Future work

The work presented in this thesis has revealed some interesting phenomena which require further investigation and the development of TVR-PCR has opened new avenues for research into the dynamics of telomeres.

Is there a gene close to the Xp:Yp telomere?

In an attempt to explain the existence of highly diverged haplotypes in the DNA flanking the Xp:Yp telomere I have invoked the possibility that some form of balancing selection is operating to maintain polymorphism at this locus. In order for selection to operate, a gene would have to be present at or near this locus. Similarity was detected between sequences only 2kb from the Xp:Yp telomere and five cDNA clones. More sequence data is required to determine whether the sequence similarity to the cDNAs extends internally and whether putative open reading frames can be detected. In addition, probes consisting of the DNA between DXYS14 and the Xp:Yp telomere could be used to screen cDNA libraries, or exon trapping could also be used to identify coding regions from this region (Buckler *et al.*, 1991).

A comprehensive screen for polymorphism at the orthologous locus in chimpanzees and gorillas could be carried out. If the positions identified as polymorphic in humans were also polymorphic in chimpanzees and gorillas then this would support the argument for the presence of a gene under the influence of balancing selection. The generation of additional sequence proximal to the human Xp:Yp telomere junction region could allow screens for polymorphism to determine how far the high frequency of sequence polymorphism and linkage disequilibrium extends internally. A scan for sequence polymorphism in the distal flanking DNA of the DXYS14 locus should be carried out to determine whether the linkage disequilibrium extends as far as this locus. It probably would not be possible to detect linkage disequilibrium at this hypervariable minisatellite locus itself (Inglehearn and Cooke, 1990), as the mutational

mechanisms generating the variation may rapidly break down any linkage disequilibrium.

Recombination rates at the very end of the chromosome. A screen for recombination between DXYS14 and the telomere by the analysis of pedigrees has been described in this thesis but none were identified. Further investigation of recombination rates in this region could be achieved by the analysis of recombination events derived from sperm DNA, as has been carried out for some of the Xp:Yp pseudoautosomal region (Schmitt *et al.*, 1994).

Applications of TVR-PCR

Forensics. The allelic diversity in the distribution of telomere and variant repeats in the Xp:Yp telomere revealed by TVR-PCR suggests that this technique could be a useful tool in forensic and parentage analysis. This would be facilitated by the development of a fluorescent based detection system for TVR-PCR such as the use of automatic DNA sequencing machines as described in this thesis. This would allow greater sample throughput and obviates the need for the use of radioactivity. The recently introduced national forensic database is based upon the PCR analysis of short tandem repeat loci utilising fluorescent detection systems on automated DNA sequencing machines, thus, automation of TVR-PCR could easily be incorporated into such a database. Realistically, TVR-PCR may be most useful in the analysis of highly degraded DNA samples, since the short repeat unit means that only a short fragment of DNA could yield a large amount of information from a single locus. The power of this analysis could be increased by including analysis of other telomeres.

At additional loci. Work is in progress in this laboratory to develop a TVR-PCR system at another single copy telomere at the end of the long arm of chromosome 7 (J. Coleman pers. comm.). This should facilitate a detailed analysis of variation at an additional chromosome end to complement the study of the Xp:Yp telomere described in this thesis. It will also help determine whether strong linkage disequilibrium is a general feature of chromosome ends or just a phenomenon of the Xp:Yp telomere. TVR-PCR is in principal applicable to any telomere provided adjacent sequence can be obtained which would allow amplification from a single telomere.

TVR-PCR could also be applied to interstitial telomere-like repeats. For example, the locus within band 2q13, contains a head-to-head array of telomere repeats and is thought to be a site of a chromosomal fusion of two ancestral ape chromosomes to form the human chromosome 2 (Ijdo *et al.*, 1991). This locus

contains telomere and variant repeats and therefore should be amenable to TVR-PCR. This would be an interesting comparison of telomere-like repeat turnover at an interstitial locus and at a telomere, different factors are likely to influence the evolution of these loci. The chromosomal fusion presumably occurred only once in the human lineage after the divergence from the chimpanzee lineage between 4.7 to 7.1Myr ago. Consequently, any variation detected at the 2q fusion point in modern humans would allow an estimate of the rate of evolution to be made with a reasonable degree of accuracy.

Telomere Dynamics. Since TVR-PCR is based upon the polymerase chain reaction it is potentially a very sensitive technique. It is hoped that the sensitivity can be increased such that it is possible to analyse telomeres from a limited number of cells. TVR-PCR could then be used to analyse the dynamics of a single telomere in senescing cells. Telomere length changes and fusions (Pathak *et al.*, 1988; Hastie *et al.*, 1990) have been observed in numerous tumours and it may be possible to use TVR-PCR to dissect such events at the molecular level.

TVR-PCR provides a sensitive system for the analysis of variation at loci with a short repeat. It may be useful to analyse changes in telomere repeat arrays in individuals with short repeat instability, such as that observed in individuals who suffer from HNPCC RER+ tumours (Thibodeau *et al.*, 1993; Ionov *et al.*, 1993). Comparison of normal and tumour DNAs from such individuals could observe in detail changes at a six bp repeat locus.

Mutation in the Xp:Yp telomere. The types of mutational mechanisms that generate the observed variation in the Xp:Yp telomere can be inferred from the comparison of groups of related alleles sampled from the population. However, the most informative way to observe the types of mutation mechanisms involved is to analyse the outcomes of *de novo* mutations. This can be done by screening for mutation in pedigrees as described in this thesis, but this is labour intensive, as even a high mutation rate requires that a large number of meioses are scanned to detect mutations. The most useful set of large kindreds for scanning for mutations are those in the CEPH panel of DNAs. However, these DNAs are derived from lymphoblastoid cell lines and consequently it can be difficult to determine whether any mutations detected are a consequence of a genuine meiotic mutation event, or a consequence of an *in vitro* somatic mutation in the lymphoblastoid cell lines (Banchs *et al.*, 1994). One unlimited source of mutations would be from DNA obtained from sperm and techniques have been developed to utilise mutations or recombination events occurring in sperm DNA (Schmitt *et al.*, 1994; Jeffreys *et al.*, 1994; Zangenberg *et al.*, 1995). Telomere repeat arrays would have to be found that contain unique sites for PCR primers to anneal. This should be

possible given the observed variation in the distribution of different repeat types, for example, some telomeres contain single blocks of TCAGGG repeats and a unique priming site would be present at the distal most junction of the block of TCAGGG repeats with TTAGGG repeats. Amplification of sperm DNA from a heterozygous individual, with an allele specific primer in the flanking DNA specific for the haplotype associated with the telomere from which the unique telomere primers was designed, might allow the amplification of mutations that resulted from intra-allelic events. In addition, amplification with a flanking primer specific for the other haplotype in the heterozygous individual and the telomere unique primer, might allow the amplification of mutations resulting from inter-allelic events. These PCR products could be detected on denaturing polyacrylamide gels and any putative mutants would be structurally characterised by TVR-PCR. SSCP analysis could be used to detect mutations resulting from events that do not involve changes in repeat number, for example, gene conversion-like events. It may also be possible using these techniques, to detect exchange events between telomeres from different chromosomes, though the rate of such events may be below the level of detection even with sensitive PCR assays.

If the techniques for the isolation of mutants from sperm DNA are successful the analysis could be taken further. Given the most likely mutational mechanism within the Xp:Yp telomere is replication slippage, then one may anticipate that the longer the block of identical repeat units the greater the probability of mutation. This assumption is based upon the observation that variation and instability at simple repeat loci is a function of the length of uninterrupted single repeat unit types (Weber, 1990; Richards and Sutherland, 1992; Eichler *et al.*, 1994). Analysis of telomeres with different sized arrays of identical repeat units would address whether the length of identical repeat unit types has an effect upon the mutation rate by replication slippage events at the Xp:Yp telomere.

Human population diversity. The variation detected in both the Xp:Yp telomere and the flanking DNA may provide a system for the analysis of the diversity of modern human populations. There are two contrasting models concerning the geographical distribution of modern human populations, namely the African origin (out of Africa) and multiregional evolution (regional continuity). The African replacement theory proposes that modern human populations are derived from migrations of *Homo sapiens* from Africa approximately 100,000 years ago, replacing pre-existing *Homo* populations (Stringer and Andrews, 1988). The Multiregional model asserts that human populations have pursued essentially parallel evolution over the last 1Myr, with

some interbreeding to maintain a similar pattern of evolutionary development (Wolpoff, 1989). The main point of difference between these two models is how and when the divergence between the different populations occurred. The African origin model suggests that the differences between modern human populations are the result of short term evolutionary processes operating over approximately 100,000 years. Whereas the multiregional hypothesis asserts that these differences are the result of evolutionary processes occurring over a period of upto 1Myr. Much evidence has come to light in favour of the African replacement theory, particularly from the analysis of DNA sequence polymorphism. For example, variation in the mtDNA D-loop sequence suggested that the sequences were too similar to be the products of independent evolution over a long period of time, also the greatest variation was observed in Africans (Stoneking and Cann, 1989). Therefore, it has been argued that the origin of modern humans was from Africa. The diversity in the African populations in the DNA flanking the Xp:Yp telomere appears to be greater than in the other human populations analysed. The African haplotype A sequence appears to be different from that of Caucasians, however as described in chapter 5 it seems likely that the Caucasian haplotype A may also be represented in the African population. This suggests that diversity in other human populations may be derived from the African population, supporting the African replacement hypothesis. It may be possible to use variation within the telomere repeat array to trace lineages, but if the rate of sequence turnover in telomeres is too great, it would be impossible to track lineages over a long period of evolutionary time. Additional sequence analysis of haplotypes in the DNA flanking the Xp:Yp telomere and a larger sample of telomeric alleles from African and other human populations would enhance the study of human population diversity.

Further characterisation of sequences from chimpanzees and gorillas orthologous to the human Xp:Yp telomere

The work presented in this thesis demonstrates that the Xp:Yp telomere is not in the same location in chimpanzees and gorillas as it is in humans. In chimpanzees, PCR analysis demonstrated that a 32bp subterminal satellite was present instead of the telomere. The same analysis in gorillas suggested that a telomere was also not present, but instead there were two small interstitial blocks of telomere-like repeats. No subterminal satellite was detected in gorillas by PCR analysis, although *in situ* hybridisation studies have revealed that it is present in the terminal region of short arms of the gorilla X and Y chromosomes. It would be of interest to obtain more information about the sequences present in chimpanzees and gorillas in place of the telomere found in humans. My preliminary data suggests that the SINE is truncated at a different site in humans,

chimpanzees and gorillas. Further investigations may give some indication as to what sequences were present in the progenitor that gave rise to the chimpanzee and human lineages and whether additional changes have occurred in the gorilla lineage.

In conclusion

I have speculated upon the mechanisms underlying the unexpected properties of the Xp:Yp telomere. Additional experimentation is required to test these hypotheses and to determine whether they are general features of sequences at the ends of human chromosomes or unique to the Xp:Yp pseudoautosomal telomere. Potentially, TVR-PCR may provide a useful system for the analysis of telomeric repeat turnover *in vitro* and *in vivo* and could provide an alternative powerful technique for individual identification.

REFERENCES

Aaltonen, L. A., Peltomaki, P., Leach, F. S., Sistonen, P., Pylkkanen, L., Mecklin, J. P., Jarvinen, H., Powell, S. M., Jen, J., Hamilton, S. R., Petersen, G. M., Kinzler, K. W., Vogelstein, B. and de la Chapelle, A. (1993). Clues to the pathogenesis of familial colorectal-cancer. *Science* **260**: 812-816.

Adamson, D. J. A., King, D. J. and Haites, N. E. (1992). Significant telomere shortening in childhood leukemia. *Cancer Genetics and Cytogenetics* **61**: 204-206.

Allen, M. J. (1994). Transgenes of the human hypervariable minisatellite MS32., Ph.D thesis, University of Leicester.

Allshire, R. C., Dempster, M. and Hastie, N. D. (1989). Human telomeres contain at least 3 types of G-rich repeat distributed non-randomly. *Nucleic Acids Res.* **17**: 4611-4627.

Allshire, R. C., Gosden, J. R., Cross, S. H., Cranston, G., Rout, D., Sugawara, N., Szostak, J. W., Fantes, P. A. and Hastie, N. D. (1988). Telomeric repeat from *T. thermophila* cross hybridizes with human telomeres. *Nature* **332**: 656-659.

Allsopp, R. C., Vaziri, H., Patterson, C., Goldstein, S., Younglai, E. V., Futcher, A. B., Greider, C. W. and Harley, C. B. (1992). Telomere length predicts replicative capacity of human fibroblasts. *Proc. Natl. Acad. Sci. USA* **89**: 10114-10118.

Allsopp, R. C. and Harley, C. B. (1995). Evidence for a critical telomere length in senescent human fibroblasts. *Experimental Cell Research* **219**: 130-136.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**: 403-410.

Aparicio, O. M., Billington, B. L. and Gottschling, D. E. (1991). Modifiers of position effect are shared between telomeric and silent mating-type loci in *Saccharomyces cerevisiae*. *Cell* **66**: 1279-1287.

Armour, J. A. L., Povey, S., Jeremiah, S. and Jeffreys, A. J. (1990). Systematic cloning of human minisatellites from ordered array charomid libraries. *Genomics* **8**: 501-512.

Armour, J. A. L., Harris, P. C. and Jeffreys, A. J. (1993). Allelic diversity at minisatellite MS205 (D16S309): evidence for polarized variability. *Hum. Mol. Genet.* **2**: 1137-1145.

Armour, J. A. L., Neumann, R., Gobert, S. and Jeffreys, A. J. (1994). Isolation of human simple repeat loci by hybridization selection. *Hum. Mol. Genet.* **3**: 599-605.

Banchs, I., Bosch, A., Guimera, J., Lazaro, C., Puig, A. and Estivill, X. (1994). New alleles at microsatellite loci in ceph families mainly arise from somatic mutations in the lymphoblastoid cell-lines. *Human Mutation* **3**: 365-372.

Bayne, R. A. L., Broccoli, D., Taggart, M. H., Thomson, E. J., Farr, C. J. and Cooke, H. J. (1994). Sandwiching of a gene within 12 kb of a functional telomere and alpha satellite does not result in silencing. *Hum. Mol. Genet.* **3**: 539-546.

Bender, M. A., Preston, R. J., Leonard, R. C., Pyatt, B. E. and Gooch, P. C. (1989). Chromosomal aberration and sister-chromatid exchange frequencies in peripheral-blood lymphocytes of a large human-population sample .2. extension of age range. *Mutation Research* **212**: 149-154.

Benn, P. A. (1976). Specific chromosome aberrations in senescent fibroblast cell lines derived from human embroyos. *Am. J. Hum. Genet.* **28**: 465-473.

Biessmann, H. and Mason, J. M. (1988). Progressive loss of DNA sequences from terminal chromosome deficiencies in *Drosophila melanogaster*. *EMBO J.* **7**: 1081-1086.

Biessmann, H., Champion, L. E., Ohair, M., Ikenaga, K., Kasravi, B. and Mason, J. M. (1992). Frequent transpositions of *Drosophila melanogaster* HeT-A transposable elements to receding chromosome ends. *EMBO J.* **11**: 4459-4469.

Biessmann, H., Valgeirsdottir, K., Lofsky, A., Chin, C., Ginther, B., Levis, R. W. and Pardue, M. L. (1992). HeT-A, a transposable element specifically involved in healing broken chromosome ends in *Drosophila melanogaster*. *Mol. Cell. Biol.* **12**: 3910-3918.

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. and Wiley, D. C. (1987). The foreign antigen-binding site and T-cell recognition regions of class-I histocompatibility antigens. *Nature* **329**: 512-518.

Blackburn, E. H. and Gall, J. G. (1978). A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in *Tetrahymena*. *J. Mol. Biol.* **120**: 33-53.

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

Bodmer, W. F. (1972). Evolutionary significance of the HLA system. *Nature* **237**: 139-145.

Britten, R. J. and Kohne, D. E. (1968). Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genome of higher organisms. *Science* **161**: 529-540.

Britten, R. J., Baron, W. F., Stout, D. B. and Davidson, E. H. (1988). Sources and evolution of human Alu repeated sequences. *Proc. Natl. Acad. Sci. USA* **85**: 4770-4774.

Broccoli, D., Miller, O. J. and Miller, D. A. (1992). Isolation and characterization of a mouse subtelomeric sequence. *Chromosoma* **101**: 442-447.

Broccoli, D., Young, J. W. and de Lange, T. (1995). Telomerase activity in normal and malignant hematopoietic-cells. *Proc. Natl. Acad. Sci. USA* **92**: 9082-9086.

Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., hunter, K., Stanton, V. P., Thirion, J. P., Hudson, T., Sohn, R., Zemelman, B., snell, R. G., Rundle, S. A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P. S., Shaw, D. J. and Housman, D. E. (1992). Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* **68**: 799-808.

Brown, W. R. A. (1988). A physical map of the human pseudoautosomal region. *EMBO J.* **7**: 2377-2385.

Brown, W. R. A. (1989). Molecular-cloning of human telomeres in yeast. *Nature* **338**: 774-776.

Brown, W. R. A., 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**: 119-132.

Bryan, T. M., Englezou, A., Gupta, J., Bacchetti, S. and Reddel, R. R. (1995). Telomere elongation in immortal human cells without detectable telomerase activity. *EMBO J.* **14**: 4240-4248.

Buchman, A. R., Kimmerly, W. J., Rine, J. and Kornberg, R. D. (1988). 2 DNA-binding factors recognize specific sequences at silencers, upstream activating sequences, autonomously replicating sequences, and telomeres in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **8**: 210-225.

Buckler, A. J., Chang, D. D., Graw, S. L., Brook, J. D., Haber, D. A., Sharp, P. A., and Housman, D. E. (1991). Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. USA* **88**: 4005-4009.

Callegarijacques, S. M., Salzano, F. M., Weimer, T. A., Hutz, M. H., Black, F. L., Santos, S., Guerreiro, J. F., Mestriner, M. A. and Pandey, J. P. (1994). Further blood genetic-studies on amazonian diversity: data from 4 indian groups. *Annals of Human Biology* **21**: 465-481.

Cardenas, M. E., Laroche, T. and Gasser, S. M. (1990). The composition and morphology of yeast nuclear scaffolds. *Journal of Cell Science* **96**: 439-450.

Carpenter, A. T. C. (1979). Recombination nodules and synaptonemal complex in recombination-defective females of *Drosophila melanogaster*. *Chromosoma* **75**: 259-292.

Carpenter, A. T. C. (1987). Gene conversion, recombination nodules, and the intiation of meiotic synapsis. *BioEssays* **6**: 232-236.

Chadeneau, C., Hay, K., Hirte, H. W., Gallinger, S. and Bacchetti, S. (1995). Telomerase activity associated with acquisition of malignancy in human colorectal-cancer. *Cancer Research* **55**: 2533-2536.

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

Chang, S. E. (1985). *In vitro* transformation of human epithelial-cells. *Biochimica et Biophysica Acta* **823**: 161-194.

Cheng, J. F., Smith, C. L. and Cantor, C. R. (1989). Isolation and characterization of a human telomere. *Nucleic Acids Res.* **17**: 6109-6127.

Cheng, J. F., Smith, C. L. and Cantor, C. R. (1991). Structural and transcriptional analysis of a human subtelomeric repeat. *Nucleic Acids Res.* **19**: 149-154.

Chikashige, Y., Ding, D. Q., Funabiki, H., Haraguchi, T., Mashiko, S., Yanagida, M. and Hiraoka, Y. (1994). Telomere-led premeiotic chromosome movement in fission yeast. *Science* **264**: 270-273.

Chong, L., van Steensel, B., Broccoli, D., Erdjument-Bromage, H., Hanish, J., Tempst, P. and de Lange, T. (1995). A human telomeric protein. *Science*. **270**: 1663-1667.

Chung, H.-M. M., Shea, C., Fields, S., Taub, R. N. and Van der Ploeg, L. H. T. (1990). Architectural organization in the interphase nucleus of the protozoan *Trypanosoma brucei*: location of telomeres and minichromosomes. *EMBO J.* **9**: 2611-2619.

Chung, M. Y., Ranum, L., Duvick, L. A., Servadio, A., Zoghbi, H. Y. and Orr, H. T. (1993). Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type-I. *Nature Genet.* 5: 254-258.

Church, G. M. and Gilbert, W. (1984). Genomic Sequencing. *Proc. Natl. Acad. Sci. USA* 81: 1991-1995.

Cohn, M. and Blackburn, E. H. (1995). Telomerase in yeast. *Science* 269: 396-400.

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

Conrad, M. N., Wright, J. H., Wolf, A. J. and Zakian, V. A. (1990). RAP1 protein interacts with yeast telomeres invivo: overproduction alters telomere structure and decreases chromosome stability. *Cell* 63: 739-750.

Cooke, H. J., Brown, W. and Rappold, G. A. (1985). Hypervariable telomeric sequences from the human sex-chromosomes are pseudoautosomal. *Nature* 317: 687-692.

Cooke, H. J. and Smith, B. A. (1986). Variability at the telomeres of the human X/Y pseudoautosomal region. *Cold Spring Harbor Symposia on Quantitative Biology* 51: 213-219.

Cooper, D. N., Smith, B. A., Cooke, H. J., Niemann, S. and Schmidtke, J. (1985). An estimate of unique DNA-sequence heterozygosity in the human genome. *Hum. Genet.* 69: 201-205.

Corcoran, L. M., Thompson, J. K., Walliker, D. and Kemp, D. J. (1988). Homologous recombination within subtelomeric repeat sequences generates chromosome size polymorphisms in *P. falciparum*. *Cell* 53: 807-813.

Counter, C. M., Avilion, A. A., Lefeuvre, C. E., Stewart, N. G., Greider, C. W., Harley, C. B. and Bacchetti, S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J.* 11: 1921-1929.

Counter, C. M., Botelho, F. M., Wang, P., Harley, C. B. and Bacchetti, S. (1994). Stabilization of short telomeres and telomerase activity accompany immortalization of epstein barr virus-transformed human B-lymphocytes. *Journal of Virology* 68: 3410-3414.

Counter, C. M., Hirte, H. W., Bacchetti, S. and Harley, C. B. (1994). Telomerase activity in human ovarian-carcinoma. *Proc. Natl. Acad. Sci. USA* 91: 2900-2904.

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: 2315-2320.

Craig, J. M. and Bickmore, W. A. (1994). The distribution of CpG islands in mammalian chromosomes. *Nature Genet.* 7: 376-382.

Cross, S. H., Allshire, R. C., McKay, S. J., McGill, N. I. and Cooke, H. J. (1989). Cloning of human telomeres by complementation in yeast. *Nature* 338: 771-774.

Cross, S., Lindsey, J., Fantes, J., McKay, S., McGill, N. and Cooke, H. (1990). The structure of a subterminal repeated sequence present on many human-chromosomes. *Nucleic Acids Res.* 18: 6649-6657.

Cullen, M., Erlich, H., Klitz, W. and Carrington, M. (1995). Molecular mapping of a recombination hotspot located in the 2nd intron of the human TAP2 locus. *Am. J. Hum. Genet.* **56**: 1350-1358.

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

de Lange, T., Shiue, L., Myers, R. M., Cox, D. R., Naylor, S. L., Killery, A. M. and Varmus, H. E. (1990). Structure and variability of human-chromosome ends. *Mol. Cell. Biol.* **10**: 518-527.

de Lange, T. (1992). Human telomeres are attached to the nuclear matrix. *EMBO J.* **11**: 717-724.

Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programmes for the VAX. *Nucleic Acids Res.* **12**: 387-395.

DiPaolo, J. A. (1983). Relative difficulties in transforming human and animal-cells *in vitro*. *Journal of the National Cancer Institute* **70**: 3-8.

Durfy, S. J. and Willard, H. F. (1990). Concerted evolution of primate alpha-satellite DNA: evidence for an ancestral sequence shared by gorilla and human X-chromosome alpha-satellite. *J. Mol. Biol.* **216**: 555-566.

Economou, E. P., Bergen, A. W., Warren, A. C. and Antonarakis, S. E. (1990). The polydeoxyadenylate tract of alu repetitive elements is polymorphic in the human genome. *Proc. Natl. Acad. Sci. USA* **87**: 2951-2954.

Eichler, E. E., Holden, J. J. A., Popovich, B. W., Reiss, A. L., Snow, K., Thibodeau, S. N., Richards, C. S., Ward, P. A. and Nelson, D. L. (1994). Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nature Genet.* **8**: 88-94.

Ellis, N. A., Goodfellow, P. J., Pym, B., Smith, M., Palmer, M., Frischauf, A. M. and Goodfellow, P. N. (1989). The pseudoautosomal boundary in man is defined by an alu repeat sequence inserted on the Y-chromosome. *Nature* **337**: 81-84.

Enssle, K. H., Wagner, H. and Fleischer, B. (1987). Human mumps virus-specific cyto-toxic lymphocytes-T: quantitative analysis of HLA restriction. *Human Immunology* **18**: 135-149.

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* **3**: 87-112.

Fajkus, J., Kovarik, A. and Bezdek, M. (1995). Organization of telomeric and subtelomeric chromatin in the higher plant *Nicotiana tabacum*. *Mol. Gen. Genet.* **247**: 633-638.

Fang, G. W. and Cech, T. R. (1993). The beta-subunit of oxytricha telomere-binding protein promotes G-quartet formation by telomeric DNA. *Cell* **74**: 875-885.

Farr, C. J., Stevanovic, M., Thomson, E. J., Goodfellow, P. N. and Cooke, H. J. (1992). Telomere-associated chromosome fragmentation: applications in genome manipulation and analysis. *Nature Genet.* **2**: 275-282.

Feng, J. L., Funk, W. D., Wang, S. S., Weinrich, S. L., Avilion, A. A., Chiu, C. P., Adams, R. R., Chang, E., Allsopp, R. C., Yu, J. H., Le, S. Y., West, M. D., Harley, C. B., Andrews, W. H., Greider, C. W. and Villeponteau, B. (1995). The RNA component of human telomerase. *Science* **269**: 1236-1241.

Fienberg, A. P. and Vogelstein, B. (1984). A technique for radiolabelling DNA restriction endonuclease fragments to a high specific activity. *Anal. Biochem.*: 266-267.

Fitzgerald, P. H. and Morris, C. M. (1984). Telomeric association in B-cell lymphoid leukemia. *Hum. Genet.* 67: 385-390.

Flint, J., Craddock, C. F., Villegas, A., Bentley, D. P., Williams, H. J., Galanello, R., Cao, A., Wood, W. G., Ayyub, H. and Higgs, D. R. (1994). Healing of broken human-chromosomes by the addition of telomeric repeats. *Am. J. Hum. Genet.* 55: 505-512.

Flint, J., Wilkie, A. O. M., Buckle, V. J., Winter, R. M., Holland, A. J. and McDermid, H. E. (1995). The detection of subtelomeric chromosomal rearrangements in idiopathic mental-retardation. *Nature Genet.* 9: 132-140.

Funabiki, H., Hagan, I., Uzawa, S. and Yanagida, M. (1993). Cell cycle-dependent specific positioning and clustering of centromeres and telomeres in fission yeast. *Journal of Cell Biology* 121: 961-976.

Freije, D., Helms, C., Watson, M. S. and Donis-Keller, D. (1992). Identification of a second pseudoautosomal region near the Xq and Yq telomeres. *Science* 258: 1784-1787.

Ganal, M. W., Laptain, N. L. V. and Tanksley, S. D. (1991). Macrostructure of the tomato telomeres. *Plant Cell* 3: 87-94.

Gilson, E., Roberge, M., Giraldo, R., Rhodes, D. and Gasser, S. M. (1993). Distortion of the DNA double helix by RAP1 at silencers and multiple telomeric binding-sites. *J. Mol. Biol.* 231: 293-310.

Gilson, E., Mullers, I., Sogo, J., Laroche, T. and Gasser, S. M. (1994). RAP1 stimulates single-strand to double-strand association of yeast telomeric DNA: implications for telomere-telomere interactions. *Nucleic Acids Res.* 22: 5310-5320.

Giraldo, R. and Rhodes, D. (1994). The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J.* 13: 2411-2420.

Giraldo, R., Suzuki, M., Chapman, L. and Rhodes, D. (1994). Promotion of parallel DNA quadruplexes by a yeast telomere binding-protein: a circular dichroism study. *Proc. Natl. Acad. Sci. USA* 91: 7658-7662.

Goldstein, S. (1990). Replicative senescence: The human fibroblast comes of age. *Science* 249: 1129-1133.

Gottschling, D. E. and Cech, T. R. (1984). Chromatin structure of the molecular ends of *Oxytricha* macronuclear DNA: Phased nucleosomes and a telomeric complex. *Cell* 38: 501-510.

Gottschling, D. E. and Zakian, V. A. (1986). Telomere proteins that recognize the molecular ends of *Oyxtricha* macronuclear DNA. *Journal of Cell Biology* 103: A289.

Gottschling, D. E. and Zakian, V. A. (1986). Telomere proteins: specific recognition and protection of the natural termini of *Oxytricha* macronuclear DNA. *Cell* 47: 195-205.

Gottschling, D. E., Aparicio, O. M., Billington, B. L. and Zakian, V. A. (1990). Position effect at *Saccharomyces cerevisiae* telomeres: reversible repression of pol II transcription. *Cell* 63: 751-762.

Gottschling, D. E. (1992). Telomere-proximal DNA in *Saccharomyces cerevisiae* is refractory to methyltransferase activity *in vivo*. *Proc. Natl. Acad. Sci. USA* **89**: 4062-4065.

Graham, I. R. and Chambers, A. (1994). Use of a selection technique to identify the diversity of binding sites for the yeast RAP1 transcription factor. *Nucleic Acids Res.* **22**: 124-130.

Greenwell, P. W., Kronmal, S. L., Porter, S. E., Gassenhuber, J., Obermaier, B. and Petes, T. D. (1995). Tel1, a gene involved in controlling telomere length in saccharomyces-cerevisiae, is homologous to the human ataxia telangiectasia gene. *Cell* **82**: 823-829.

Greider, C. W. and Blackburn, E. H. (1985). Identification of a specific telomere terminal transferase-activity in tetrahymena extracts. *Cell* **43**: 405-413.

Greider, C. W. and Blackburn, E. H. (1989). A telomeric sequence in the RNA of tetrahymena telomerase required for telomere repeat synthesis. *Nature* **337**: 331-337.

Greig, G. M. and Willard, H. F. (1992). Beta-satellite DNA characterization and localization of 2 subfamilies from the distal and proximal short arms of the human acrocentric chromosomes. *Genomics* **12**: 573-580.

Greig, G. M., Warburton, P. E. and Willard, H. F. (1993). Organization and evolution of an alpha-satellite DNA subset shared by human chromosome-13 and chromosome-21. *Journal of Molecular Evolution* **37**: 464-475.

Guerrini, A. M., Camponeschi, B., Ascenzioni, F., Piccolella, E. and Donini, P. (1993). Subtelomeric as well as telomeric sequences are lost from chromosomes in proliferating B-lymphocytes. *Hum. Mol. Genet.* **2**: 455-460.

Gyllensten, U. B. and Erlich, H. A. (1988). Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proc. Natl. Acad. Sci. (USA)* **85**: 7652-7656.

Gyllensten, U. B. and Erlich, H. A. (1989). Ancient roots for polymorphism at the HLA-DQα locus in primates. *Proc. Natl. Acad. Sci. USA*: 9986-9990.

Hardin, C. C., Henderson, E., Watson, T. and Prosser, J. K. (1991). Monovalent cation induced structural transitions in telomeric DNAs: G-DNA folding intermediates. *Biochemistry* **30**: 4460-4472.

Hardy, C., Sussel, L. and Shore, D. (1992). A rap1-interacting protein involved in transcriptional silencing and telomere length regulation. *Genes & Development* **6**: 801-814.

Harley, C. B., Futcher, A. B. and Greider, C. W. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature* **345**: 458-460.

Harley, C. B. (1991). Telomere loss: mitotic clock or genetic time bomb. *Mutation Research* **256**: 271-282.

Harrington, L. A. and Greider, C. W. (1991). Telomerase primer specificity and chromosome healing. *Nature* **353**: 451-454.

Harris, P. C. and Thomas, S. (1992). Length polymorphism of the subtelomeric regions of both the short (p) and long arms (q) of chromosome16. *Cytogenetics and Cell Genetics* **60**: 171-172.

Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D. K. and Allshire, R. C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. *Nature* **346**: 866-868.

Henderson, E. R. and Blackburn, E. H. (1989). An overhanging 3' terminus is a conserved feature of telomeres. *Mol. Cell. Biol.* **9**: 345-348.

Henderson, S. T. and Petes, T. D. (1992). Instability of simple squence DNA in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 12: 2749-2757.

Henikoff, S. (1990). Position effect varigation after 60 years. *Trends in Genetics* **6**: 422-426

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M. and Wilson, R. (1995). The WashU-Merk EST project. Washington University School of Medicine, 4444 Forest park parkway, Box 8501, St. Louis, MO 6310.

Hiyama, K., Hirai, Y., Kyoizumi, S., Akiyama, M., Hiyama, E., Piatyszek, M. A., Shay, J. W., Ishioka, S. and Yamakido, M. (1995). Activation of telomerase in human-lymphocytes and hematopoietic progenitor cells. *Journal of Immunology* **155**: 3711-3715.

Hoffolsen, P., Meling, G. I. and Olaisen, B. (1995). Somatic mutations in vntr-locus D1S7 in human colorectal carcinomas are associated with microsatellite instability. *Human Mutation* **5**: 329-332.

Hoglund, M., Mitelman, F. and Mandahl, N. (1995). A human 12p-derived cosmid hybridizing to subsets of human and chimpanzee telomeres. *Cytogenetics and Cell Genetics* **70**: 88-91.

Horowitz, H., Thorburn, P. and Haboer, J. E. (1984). Rearrangements of highly polymorphic regions near telomeres of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4**: 2509-2517.

Hughes, A. L. and Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**: 167-170.

Hulten, M. (1974). Chiasma distribution at diakinesis in the normal human male. *Hereditas* **76**: 55-78.

Hunt, P. A. and LeMaire, R. (1992). Sex-chromosome pairing: evidence that the behavior of the pseudoautosomal region differs during male and female meiosis. *Am. J. Hum. Genet.* **50**: 1162-1170.

Ijdo, J. W., Baldini, A., Ward, D. C., Reeders, S. T. and Wells, R. A. (1991). Origin of human chromosome 2: An ancestral telomere-telomere fusion. *Proc. Natl. Acad. Sci. USA* **88**: 9051-9055.

Ijdo, J. W., Lindsay, E. A., Wells, R. A. and Baldini, A. (1992). Multiple variants in subtelomeric regions of normal karyotypes. *Genomics* 14: 1019-1025.

Inglehearn, C. F. and Cooke, H. J. (1990). A VNTR immediately adjacent to the human pseudoautosomal telomere. *Nucleic Acids Res.* **18**: 471-476.

Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D. and Perucho, M. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* **363**: 558-561.

Ishikawa, F., Matunis, M. J., Dreyfuss, G. and Cech, T. R. (1993). Nuclear proteins that bind the premessenger RNA 3' splice-site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)n. *Mol. Cell. Biol.* 13: 4301-4310.

Itzhaki, J. E., Barnett, M. A., Maccarthy, A. B., Buckle, V. J., Brown, W. R. A. and Porter, A. C. G. (1992). Targeted breakage of a human chromosome mediated by cloned human telomeric DNA. *Nature Genet.* 2: 283-287.

Jager, D. and Philippsen, P. (1989). Stabilization of dicentric chromosomes in saccharomyces-cerevisiae by telomere addition to broken ends or by centromere deletion. *EMBO J.* 8: 247-254.

Jeffreys, A. J. (1987). Highly variable minisatellites and DNA fingerprints. *Biochem. Soc. Trans.* 15: 309-317.

Jeffreys, A. J., Neumann, R. and Wilson, V. (1990). Repeat unit sequence variation in minisatellites - a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. *Cell* 60: 473-485.

Jeffreys, A. J., MacLeod, A., Tamaki, K., Neil, D. L. and Monckton, D. G. (1991). Minisatellite repeat coding as a digital approach to DNA typing. *Nature* 354: 204-209.

Jeffreys, A. J., Tamaki, K., Macleod, A., Monckton, D. G., Neil, D. L. and Armour, J. (1994). Complex gene conversion events in germline mutation at human minisatellites. *Nature Genet.* 6: 136-145.

John, B., King, M., Schweizer, D. and Mendelak, M. (1985). Equilocality of heterochromatin distribution and heterochromatin heterogeneity in acridid grasshoppers. *Chromosoma* 91: 185-200.

Jones, K. W. and Corneo, G. (1971). Location of satellite and homogenous DNA sequences on human chromosomes. *Nature New Biology* 233: 268-271.

Karpen, G. H. and Spradling, A. C. (1992). Analysis of subtelomeric heterochromatin in the drosophila minichromosome dp1187 by single P element insertional mutagenesis. *Genetics* 132: 737-753.

Kennedy, B. K., Austriaco, N. R., Zhang, J. S. and Guarente, L. (1995). Mutation in the silencing gene SIR4 can delay aging in *Saccharomyces cerevisiae*. *Cell* 80: 485-496.

Kidd, J. R., Black, F. L., Weiss, K. M., Balazs, I. and Kidd, K. K. (1991). Studies of 3 amerindian populations using nuclear-DNA polymorphisms. *Human Biology* 63: 775-794.

Kilian, A. and Kleinhofs, A. (1992). Cloning and mapping of telomere-associated sequences from *Hordeum-vulgare* L. *Molecular & General Genetics* 235: 153-156.

Kim, N. W., Piatyszek, M. A., Prowse, K. R., Harley, C. B., West, M. D., Ho, P., 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: 2011-2015.

Kipling, D. and Cooke, H. J. (1990). Hypervariable ultra-long telomeres in mice. *Nature* 347: 400-402.

Kipling, D. (1995). Telomerase immortality enzyme or oncogene. *Nature Genet.* 9: 104-106.

Kit, S. (1961). Equilibrium sedimentations in density gradients of DNA preparations from animal tissues. *J. Mol. Biol.* 3: 711-716

Klobutcher, L. A., Swanton, M. T., Donini, P. and Prescott, D. M. (1981). All gene-sized DNA molecules in 4 species of hypotrichs have the same terminal sequence and an unusual 3' terminus. *Proc. Natl. Acad. Sci. USA* 78: 3015-3019.

Knoth, K., Roberds, S., Poteet, C. and Tamkun, M. (1988). Higly degenerate, inosine-containing primers specifically amplify rare cDNA using the polymerase chain reaction. *Nucleic Acids Res.* 16: 10932.

Kojis, T. L., Gatti, R. A. and Sparkes, R. S. (1991). The cytogenetics of ataxia telangiectasia. *Cancer Genet. Cytogenet.* 56: 143-158.

Kornberg, R. D. and Lorch, Y. (1992). Chromatin structure and transcription. *Annual Review of Cell Biology* 8: 563-687.

Kramer, K. M. and Haber, J. E. (1993). New telomeres in yeast are initiated with a highly selected subset of TG(1-3) repeats. *Genes & Development* 7: 2345-2356.

Kyrion, G., Liu, K., Liu, C. and Lustig, A. J. (1993). RAP1 and telomere structure regulate telomere position effects in *Saccharomyces cerevisiae*. *Genes & Development* 7: 1146-1159.

La Mantia, G., Pengue, G., Maglione, D., Pannuti, A., Pascucci, A. and Lania, L. (1989). Identification of new human repetitive sequences: characterization of the corresponding cDNAs and their expression in embryonal carcinoma-cells. *Nucleic Acids Res.* 17: 5913-5922.

Lamb, J., Harris, P. C., Wilkie, A. O. M., Wood, W. G., Dauwerse, J. H. G. and Higgs, D. R. (1993). *De novo* truncation of chromosome-16p and healing with $(TTAGGG)_n$ in the alpha-thalassemia mental-retardation syndrome (ATR-16). *Am. J. Hum. Genet.* 52: 668-676.

Larson, D. D., Spangler, E. A. and Blackburn, E. H. (1987). Dynamics of telomere length variation in *Tetrahymena thermophila*. *Cell* 50: 477-483.

Laurie, D. A. and Hulten, M. A. (1985). Further-studies on chiasma distribution and interference in the human male. *Ann. Hum. Genet.* 49: 203-214.

Lee, M. S., Gallagher, R. C., Bradley, J. and Blackburn, E. H. (1993). *In vivo* and *in vitro* studies of telomeres and telomerase. *Cold Spring Harbor Symposia on Quantitative Biology* 58: 707-718.

Levinson, G. and Gutman, G. A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4: 203-221.

Levis, R. W. (1989). Viable deletions of a telomere from a *Drosophila* chromosome. *Cell* 58: 791-801.

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: 1083-1093.

Levy, M. Z., Allsopp, R. C., Futcher, A. B., Greider, C. W. and Harley, C. B. (1992). Telomere end-replication problem and cell aging. *J. Mol. Biol.* 225: 951-960.

Lin, J. J. and Zakian, V. A. (1995). An *in vitro* assay for *Saccharomyces* telomerase requires EST1. *Cell* 81: 1127-1135.

Lindsay, S. and Bird, A. P. (1987). Use of restriction enzymes to detect potential gene sequences in mammalian DNA. *Nature* **327**: 336-338.

Lingner, J., Cooper, J. P. and Cech, T. R. (1995). Telomerase and DNA end replication: no longer a lagging strand problem? *Science* **269**: 1533-1534.

Litt, M., Kramer, P., Kort, E., Fain, P., Cox, S., Root, D., White, R., Weissenbach, J., Doniskeller, H., Gatti, R., Weber, J., Nakamura, Y., Julier, C., Hayashi, K., Spurr, N., Dean, M., Mandel, J., Kidd, K., Kruse, T., Retief, A., Bale, A., Meo, T., Vergnaud, G., Warren, S. and Willard, H. F. (1995). The CEPH consortium linkage map of human-chromosome-11. *Genomics* **27**: 101-112.

Loidl, J. (1979). C-band proximity of chiasmata and absence of terminalisation in meiotic chromosomes of yeast. *Chromosoma* **73**: 45-51.

Louis, E. J. and Haber, J. E. (1990). Mitotic recombination among subtelomeric Y' repeats in *Saccharomyces cerevisiae*. *Genetics* **124**: 547-559.

Louis, E. J. and Haber, J. E. (1990). The subtelomeric Y' repeat family in *Saccharomyces cerevisiae*: an experimental system for repeated sequence evolution. *Genetics* **124**: 533-545.

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

Luke, S. and Verma, R. S. (1993). Telomeric repeat [TTAGGG]n sequences of human-chromosomes are conserved in chimpanzee (*Pan troglodytes*). *Molecular & General Genetics* **237**: 460-462.

Lundblad, V. and Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* **57**: 633-643.

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

Lustig, A. J. (1992). Hoogsteen G-G base-pairing is dispensable for telomere healing in yeast. *Nucleic Acids Res.* **20**: 3021-3028.

Macieira-Coelho, A. and Azzarone, B. (1988). The transition from primary culture to spontaneous immortalization in mouse fibroblast populations. *Anticancer Research* **8**: 669-676.

Macina, R. A., Negorev, D. G., Spais, C., Ruthig, L. A., Hu, X. L. and Riethman, H. C. (1994). Sequence organization of the human chromosome 2q telomere. *Hum. Mol. Genet.* **3**: 1847-1853.

Mahtani, M. M. and Willard, H. F. (1990). Pulsed-field gel analysis of alpha satellite DNA at the human X chromosome centrmere: High frequency polymorphisms and array size estimate. *Genomics* **7**: 607-613.

Makarov, V. L., Lejnine, S., Bedoyan, J. and Langmore, J. P. (1993). Nucleosomal organization of telomere-specific chromatin in rat. *Cell* **73**: 775-787.

Manuelidis, L. (1978). Chromosomal location of complex and simple repeated human DNAs. *Chromosoma* **66**: 23-32.

Martin, F. H. and Castro, M. M. (1985). Base pairing involving deoxyinosine: implications for probe design. *Nucleic Acids Res.* **13**: 8927-8938.

McClintock, B. (1938). The fusion of broken ends of sister half-chromatids following chromatid breakage at meiotic anaphase. *Res. Bull.-Mo. Agric. Exp. Stn.* **290**: 1-48.

McClintock, B. (1939). The behavior of successive nuclear divisions of a chromosome broken at meiosis. *Proc. Natl. Acad. Sci. USA* **25**: 405-416.

McEachern, M. J. and Hicks, J. B. (1993). Unusally large telomeric repeats in the yeast *Candida albicans. Mol, Cell. Biol.* **13**: 551-560.

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. USA.* **91**: 3453-3457.

McKay, S. J. and Cooke, H. (1992). hnRNP A2/B1 binds specifically to single stranded vertebrate telomeric repeat $TTAGGG_n$. *Nucleic Acids Res.* **20**: 6461-6464.

McKay, S. J. and Cooke, H. (1992). A protein which specifically binds to single stranded $TTAGGG_n$ repeats. *Nucleic Acids Res.* **20**: 1387-1391.

Mehle, C., Ljungberg, B. and Roos, G. (1994). Telomere shortening in renal-cell carcinoma. *Cancer Research* **54**: 236-241.

Meyne, J., Ratliff, R. L. and Moyzis, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)n among vertebrates. *Proc. Natl. Acad. Sci. USA* **86**: 7049-7053.

Meyne, J., Hirai, H. and Imai, H. T. (1995). Fish analysis of the telomere sequences of bulldog ants (*Myrmecia formicidae*). *Chromosoma* **104**: 14-18.

Miesfeld, R., Krystal, M. and Arnheim, N. (1981). A member of a new repeated sequence family which is conserved throughout eukaryotic evolution is found between the human delta-globin and beta-globin genes. *Nucleic Acids Res.* **9**: 5931-5947.

Mirsky, A. E. and Ris, H. (1951). The deoxyribonucleic acid content of animal cells and its evolutionary significance. *J. Gen. Pysiol.* **34**: 451-462.

Miyamoto, M. M., Koop, B. F., Slightom, J. L., Goodman, M. and Tennant, M. R. (1988). Molecular systematics of higher primates: genealogical relations and classification. *Proc. Natl. Acad. Sci. USA* **85**: 7627-7631.

Mohammad-Ali, K., Eladari, M. E. and Galibert, F. (1995). Gorilla and Orangutan c-myc nucleotide sequences: Inference on hominoid phylogeny. *J. Mol. Evol.* **41**: 262-276.

Monckton, D. G., Neumann, R., Guram, T., Fretwell, N., Tamaki, K., Macleod, A. and Jeffreys, A. J. (1994). Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism. *Nature Genet.* **8**: 162-170.

Morin, G. B. (1989). The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell* **59**: 521-529.

Morin, G. B. (1991). Recognition of a chromosome truncation site associated with alpha thalassemia by human telomerase. *Nature* **353**: 454-456.

Morton, N. E. (1982). *Outline of Genetic Epidemilogy*, Basel, Karger.

Moyzis, R. K., Buckingham, J. M., Cram, L. S., Dani, M., Deaven, L. L., Jones, M. D., Meyne, J., Ratliff, R. L. and Wu, J. R. (1988). A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci. USA* 85: 6622-6626.

Muller, F., Wicky, C., Spicher, A. and Tobler, H. (1991). New telomere formation after developmentally regulated chromosomal breakage during the process of chromatin diminution in *Ascaris lumbricoides*. *Cell* 67: 815-822.

Muller, H. J. (1938). The remaking of chromosomes. *Collect. Net* 8: 182-195.

Muller, H. J. (1940). An analysis of the process of structural change in chromosomes of *Drosphila J. Genet.* 40: 1-66.

Murchie, A. I. H. and Lilley, D. M. J. (1994). Tetraplex folding of telomere sequences and the inclusion of adenine bases. *EMBO J.* 13: 993-1001.

Murnane, J. P., Sabatier, L., Marder, B. A. and Morgan, W. F. (1994). Telomere dynamics in an immortal human cell-line. *EMBO J.* 13: 4953-4962.

Murphy, G. L., Connell, T. D., Barritt, D. S., Koomey, M. and Cannon, J. G. (1989). Phase variation of gonococcal protein-II - regulation of gene expression by slipped-strand mispairing of a repetitive DNA-sequence. *Cell* 56: 539-547.

Murray, A. W. and Szostak, J. W. (1986). Construction and behavior of circularly permuted and telocentric chromosomes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 6: 3166-3172.

Nakamura, Y., Julier, C., Wolff, R., Holm, T., Oconnell, P., Leppert, M. and White, R. (1987). Characterization of a human midisatellite sequence. *Nucleic Acids Res.* 15: 2537-2547.

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York, Columbia University Press.

Nei, M. (1988). Relative roles of mutation and selection in the maintenance of genetic variability. *Phil, Trans. R. Soc. Lond. B* 319: 615-629.

Neil, D. L. and Jeffreys, A. J. (1993). Digital DNA typing at a 2nd hypervariable locus by minisatellite variant repeat mapping. *Hum. Mol. Genet.* 2: 1129-1135.

Newton, C. R., Graham, A., Heptinstall, L. E., Powell, S. J., Summers, C., Kalsheker, N., Smith, J. C. and Markham, A. F. (1989). Analysis of any point mutation in DNA: the amplification refractory mutation system (ARMS). *Nucl. Acids Res.* 17: 2503-2516.

NIH/CEPH, Collaborative, Mapping Group. (1992). A comprehensive genetic-linkage map of the human genome. *Science* 258: 148-162.

Norwood, H. T., Smith, J. R. and Stein, G. H. (1990). The human fibroblast-like cell model. *Handbook of the biology of aging*. San Diego, Academic Press: 131-156.

Ohta, T. (1982). Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proc. Natl. Acad. Sci. USA* 79: 1940-1944.

Okazaki, S., Tsuchida, K., Maekawa, H., Ishikawa, H. and Fujiwara, H. (1993). Identification of a pentanucleotide telomeric sequence, (TTAGG)n, in the silkworm *Bombyx mori* and in other insects. *Mol. Cell. Biol.* 13: 1424-1432.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. and Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single stranded conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* **86**: 2766-2770.

Page, D. C., Bieker, K., Brown, L. G., Hinton, S., Leppert, M., Lalouel, J. M., Lathrop, M., Nystrom-Lahti, M., de la Chapelle, A. and White, R. (1987). Linkage, physical mapping, and DNA sequence analysis of pseudoautosomal loci on the human X and Y chromosomes. *Genomics* **1**: 243-256.

Park, E.-C. and Szostak, J. W. (1990). Point mutations in the yeast histone H4 gene prevent silencing of the silent mating type locus HML. *Mol. Cell. Biol.* **10**: 4932-4934.

Parsons, R., Li, G. M., Longley, M. J., Fang, W. H., Papadopoulos, N., Jen, J., de la Chapelle, A., Kinzler, K. W., Vogelstein, B. and Modrich, P. (1993). Hypermutability and mismatch repair deficiency in RER+ tumor-cells. *Cell* **75**: 1227-1236.

Pathak, S., Wang, Z., Dhaliwal, M. K. and Sacks, P. C. (1988). Telomeric association: another characteristic of cancer chromosomes. *Cytogenetics and Cell Genetics* **47**: 227-229.

Patil, R. V. and Dekker, E. E. (1990). PCR amplification of an *Escherichia coli* gene using mixed primers containing deoxyinosine at ambigous positions in degenerate amino acid codons. *Nucleic Acids Res.* **18**: 3080.

Perrin-Pecontal, P., Gouy, M., Nigon, V. M. and Trabuchet, G. (1992). Evolution of the primate beta-globin gene region: nucleotide sequence of the delta-beta-globin intergenic region of gorilla and phylogenetic-relationships between african apes and man. *Journal of Molecular Evolution* **34**: 17-30.

Pilch, D. S., Plum, G. E. and Breslauer, K. J. (1995). The thermodynamics of DNA structures that contain lesions or guanine tetrads. *Current Opinion in Structural Biology* **5**: 334-342.

Pluta, A. F. and Zakian, V. A. (1989). Recombination occurs during telomere formation in yeast. *Nature* **337**: 429-433.

Ponzi, M., Pace, T., Dore, E. and Frontali, C. (1985). Identification of a telomeric DNA sequence in *Plasmodium berghei. EMBO J.* **4**: 2991-2996.

Price, C. M. (1990). Telomere structure in *Euplotes crassus* characterization of DNA- protein interactions and isolation of a telomere-binding protein. *Mol. Cell. Biol.* **10**: 3421-3431.

Price, C. M., Skopp, R., Krueger, J. and Williams, D. (1992). DNA recognition and binding by the *Euplotes* telomere protein. *Biochemistry* **31**: 10835-10843.

Price, C. M., Adams, A. K. and Vermeesch, J. R. (1994). Accumulation of telomerase RNA and telomere protein transcripts during telomere synthesis in *Euplotes*. *Journal of Eukaryotic Microbiology* **41**: 267-275.

Prowse, K. R., Avilion, A. A. and Greider, C. W. (1993). Identification of a nonprocessive telomerase activity from mouse cells. *Proc. Natl. Acad. Sci. USA* **90**: 1493-1497.

Prowse, K. R. and Greider, C. W. (1995). Developmental and tissue-specific regulation of mouse telomerase and telomere length. *Proc. Natl. Acad. Sci. USA* **92**: 4818-4822.

Raghuraman, M. K. and Cech, T. R. (1990). Effect of monovalent cation-induced telomeric DNA structure on the binding of *Oxytricha* telomeric protein. *Nucleic Acids Res.* **18**: 4543-4551.

Rawlins, D. J., Highett, M. I. and Shaw, P. J. (1991). Localization of telomeres in plant interphase nuclei by *in situ* hybridization and 3D confocal microscopy. *Chromosoma* **100**: 424-431.

Reed, P. W., Davies, J. L., Copeman, J. B., Bennett, S. T., Palmer, S. M., Pritchard, L. E., gough, S., Kawaguchi, Y., Cordell, H. J., Balfour, K. M., Jenkins, S. C., Powell, E. E., Vignal, A. and Todd, J. A. (1994). Chromosome-specific microsatellite sets for fluorescence-based, semiautomated genome mapping. *Nature Genet.* **7**: 390-395.

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 & Development* **7**: 1133-1145.

Richards, E. J. and Ausubel, F. M. (1988). Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell* **53**: 127-136.

Richards, R. I. and Sutherland, G. R. (1992). Dynamic mutations: a new class of mutations causing human disease. *Cell* **70**: 709-712.

Richards, R. I. and Sutherland, G. R. (1994). Simple repeat DNA is not replicated simply. *Nature Genet.* **6**: 114-116.

Riethman, H. C., Moyzis, R. K., Meyne, J., Burke, D. T. and Olson, M. V. (1989). Cloning human telomeric DNA fragments into saccharomyces-cerevisiae using a yeast-artificial-chromosome vector. *Proc. Natl. Acad. Sci. USA* **86**: 6240-6244.

Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, S., Smith, J. C. and Markham, A. F. (1990). A novel method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.* **18**: 1081.

Romero, D. P. and Blackburn, E. H. (1991). A conserved secondary structure for telomerase RNA. *Cell* **67**: 343-353.

Rouyer, F., Simmler, M. C., Johnsson, C., Vergnaud, G., Cooke, H. J. and Weissenbach, J. (1986). A gradient of sex linkage in the pseudoautosomal region of the human sex-chromosomes. *Nature* **319**: 291-295.

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**: 352-360.

Royle, N. J., Hill, M. C. and Jeffreys, A. J. (1992). Isolation of telomere junction fragments by anchored polymerase chain-reaction. *Proc. R. Soc. Lond. B* **247**: 57-61.

Royle, N. J., Armour, J. A. L., Crosier, M. and Jeffreys, A. J. (1993). Abnormal segregation of alleles in ceph pedigree DNAs arising from allele loss in lymphoblastoid DNA. *Genomics* **15**: 119-122.

Royle, N. J., Baird, D. M. and Jeffreys, A. J. (1994). A subterminal satellite located adjacent to telomeres in chimpanzees is absent from the human genome. *Nature Genet.* **6**: 52-56.

Royle, N. J. (1995). The Proterminal Regions and Telomeres of Human Chromosomes. *Advances in Genetics*, Academic Press. Inc. **32**: 273-315.

Runge, K. W. and Zakian, V. A. (1989). Introduction of extra telomeric DNA-sequences into *Saccharomyces cerevisiae* results in telomere elongation. *Mol. Cell. Biol.* **9**: 1488-1497.

Saccone, S., Desario, A., Dellavalle, G. and Bernardi, G. (1992). The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes. *Proc. Natl. Acad. Sci. USA* **89**: 4913-4917.

Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1986). Primer directed amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-491.

Sakoyama, Y., Hong, K. J., Byun, S. M., Hisajima, H., Ueda, S., Yaoita, Y., Hayashida, H., Miyata, T. and Honjo, T. (1987). Nucleotide-sequences of immunoglobulin-ε genes of chimpanzee and orangutan: DNA molecular clock and hominoid evolution. *Proc. Natl. Acad. Sci. USA* **84**: 1080-1084.

Sandell, L. L. and Zakian, V. A. (1992). Telomeric position effect in yeast. *Trends in Cell Biology* **2**: 10-14.

Savitsky, K., Barshira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., Tagle, D. A., Smith, S., Uziel, T., Sfez, S., Ashkenazi, M., Pecker, I., Frydman, M., Harnik, R., Patanjali, S. R., Simmons, A., Clines, G. A., Sartiel, A., Gatti, R. A., Chessa, L., Sanal, O., Lavin, M. F., Jaspers, N. G. J., Malcolm, A., Taylor, R., Arlett, C. F., Miki, T., Weissman, S. M., Lovett, M., Collins, F. S. and Shiloh, Y. (1995). A single Ataxia Telangiectasia gene with a product similar to PI-3 kinase. *Science* **268**: 1749-1753.

Sawyer, J. R., Goosen, L. S., Stine, K. C. and Thomas, J. R. (1994). Telomere fusion as a mechanism for the progressive loss of the short arm of chromosome-11 in an anaplastic wilms-turner. *Cancer* **74**: 767-773.

Schlotterer, C. and Tautz, D. (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* **20**: 211-215.

Schlotterer, C. and Tautz, D. (1994). Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Current Biology* **4**: 777-783.

Schmid, C. W. and Jelinek, W. R. (1982). The Alu family of dispersed repetitive sequences. *Science* **216**: 1065-1070.

Schmitt, H., Blin, N., Zankl, H. and Scherthan, H. (1994). Telomere length variation in normal and malignant human tissues. *Genes Chromosomes & Cancer* **11**: 171-177.

Schmitt, K., Lazzeroni, L. C., Foote, S., Vollrath, D., Fisher, E. M. C., Goradia, T. M., Lange, K., page, D. C. and Arnheim, N. (1994). Multipoint linkage map of the human pseudoautosomal region, based on single-sperm typing: do double crossovers occur during male meiosis. *Am. J. Hum. Genet.* **55**: 423-430.

Schwartz, H. S., Dahir, G. A. and Butler, M. G. (1993). Telomere reduction in giant-cell tumor of bone and with aging. *Cancer Genetics and Cytogenetics* **71**: 132-138.

Shampay, J., Szostak, J. W. and Blackburn, E. H. (1984). DNA sequences of telomeres maintained in yeast. *Nature* **310**: 154-157.

Shapiro, L. J., Mohandas, T., Yen, P., Speed, R. and Chandley, A. (1989). the pseudoautosomal region of the X chromosome is necessary for sex chromosome pairing. *Am. J. Hum. Genet.* **45** [**Suppl**]: A107.

Sherwood, S. W., Rush, D., Ellsworth, J. L. and Schimke, R. T. (1988). Defining cellular senescence in IMR-90 cells: a flow cytometric analysis. *Proc. Natl. Acad. Sci. USA* **85**: 9086-9090.

Shibata, D., Peinado, M. A., Ionov, Y., Malkhosyan, S. and Perucho, M. (1994). Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. *Nature Genet.* **6**: 273-281.

Shippen-Lentz, D. and Blackburn, E. H. (1989). Telomere terminal transferase activity from Euplotes crassus adds large numbers of TTTTGGGG repeats onto telomeric primers. *Mol. Cell. Biol.* **9**: 2761-2764.

Shippen-Lentz, D. and Blackburn, E. H. (1990). Functional evidence for an RNA template in telomerase. *Science* **247**: 546-552.

Singer, M. F. (1982). SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. *Cell* **28**: 433-434.

Singer, M. S. and Gottschling, D. E. (1994). TLC1: Template RNA component of *Saccharomyces cerevisiae* telomerase. *Science* **266**: 404-409.

Slade, R. W. and McCallum, H. I. (1992). Overdominant vs. frequency-dependent selection at MHC loci. *Genetics* **132**: 861-862.

Slagboom, P. E., Droog, S. and Boomsma, D. I. (1994). Genetic determination of telomere size in humans: a twin study of 3 age-groups. *Am. J. Hum. Genet.* **55**: 876-882.

Smith, G. P. (1973). Unequal crossover and the evolution of multigene families. *Cold Spring Harbor Symposia on Quantitative Biology* **38**: 507-513.

Soriano, P., Keitges, E. A., Schorderet, D. F., Harbers, K., Gartler, S. M. and Jaenisch, R. (1987). High-rate of recombination and double crossovers in the mouse pseudoautosomal region during male meiosis. *Proc. Natl. Acad. Sci. USA* **84**: 7218-7220.

Speed, R. M. and Chandley, A. C. (1990). Prophase of meiosis in human spermatocytes analysed by EM microspreading in infertile men and their controls and comparisons with human oocytes. *Hum. Genet.* **84**: 547-554.

Starling, J. A., Maule, J., Hastie, N. D. and Allshire, R. C. (1990). Extensive telomere repeat arrays in mouse are hypervariable. *Nucleic Acids Res.* **18**: 6881-6888.

Stoneking, M. and Cann, R. L. (1989). African origin of human mitochrondrial DNA. *The human revolution: behavioural and biological perspectives on the origins of modern humans.* Edinburgh, Edinburgh University Press. **1**: 17-30.

Strand, M., Prolla, T. A., Liskay, R. M. and Petes, T. D. (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**: 274-276.

Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. and Inoyne, M. (1966). Frameshift mutations and the genetic code. *Cold Spring Harbor Symposia on Quantitative Biology* **31**: 77-84.

Stringer, C. B. and Andrews, P. (1988). Genetic and fossil evidence for the origin of modern humans. *Science* **239**: 1263-1268.

Thibodeau, S. N., Bren, G. and Schaid, D. (1993). Microsatellite instability in cancer of the proximal colon. *Science* **260**: 816-819.

Thommes, P. and Hubscher, U. (1990). Eukaryotic DNA replication: enzymes and proteins acting at the fork. *European Journal of Biochemistry* **194**: 699-712.

Thomson, G. (1977). The effect of a selected locus on linked neutral loci. *Genetics* **85**: 753-788.

Thomson, G. and Klitz, W. (1987). Disequilibrium Pattern Analysis. I. Theory. *Genetics* **116**: 623-632.

Tommerup, H., Dousmanis, A. and de Lange, T. (1994). Unusual chromatin in human telomeres. *Mol. Cell. Biol.* **14**: 5777-5785.

Traverse, K. L. and Pardue, M. L. (1989). Studies of He-T DNA sequences in the pericentric regions of *Drosophila* chromosomes. *Chromosoma* **97**: 261-271.

Tyler-Smith, C. and Brown, W. R. A. (1987). Structure of the major block of alphoid satellite DNA on the human Y-chromosome. *J. Mol. Biol.* **195**: 457-470.

van Ommen, G.-J. B. (1994). Making ends meet. *Nature Genet.* **7**: 347-348.

Vaziri, H., Schachter, F., Uchida, I., Wei, L., Zhu, X. M., Effros, R., Cohen, D. and Harley, C. B. (1993). Loss of telomeric DNA during aging of normal and Trisomy 21 human lymphocytes. *Am. J. Hum. Genet.* **52**: 661-667.

Vergnaud, G. (1994). No increase in female recombination frequency in the distal part of the human pseudoautosomal region. *Genomics* **24**: 610-612.

Vermeesch, J. R., Williams, D. and Price, C. M. (1993). Telomere processing in *Euplotes*. *Nucleic Acids Res.* **21**: 5366-5371.

Vermeesch, J. R. and Price, C. M. (1994). Telomeric DNA-sequence and structure following *de novo* telomere synthesis in *Euplotes crassus*. *Mol. Cell. Biol.* **14**: 554-566.

Vourc'h, C., Taruscio, D., Boyle, A. L. and Ward, D. C. (1993). Cell cycle-dependent distribution of telomeres, centromeres, and chromosome-specific subsatellite domains in the interphase nucleus of mouse lymphocytes. *Experimental Cell Research* **205**: 142-151.

Wainwright, L. J., Middleton, P. G. and Rees, J. L. (1995). Changes in mean telomere length in basal-cell carcinomas of the skin. *Genes Chromosomes & Cancer* **12**: 45-49.

Wallace, B. M. N. and Hulten, M. A. (1985). Meiotic chromosome pairing in the normal human female. *Ann. Hum. Genet.* **49**: 215-226.

Walmsley, R. W., Chan, C. S. M., Tye, B. K. and Petes, T. D. (1984). Unusual DNA-sequences associated with the ends of yeast chromosomes. *Nature* **310**: 157-160.

Wang, K. Y., Swaminathan, S. and Bolton, P. H. (1994). Tertiary structure motif of *Oxytricha* telomere DNA. *Biochemistry* **33**: 7517-7527.

Wang, S. S. and Zakian, V. A. (1990). Telomere telomere recombination provides an express pathway for telomere acquisition. *Nature* **345**: 456-458.

Warburton, P. E., Waye, J. S. and Willard, H. F. (1993). Nonrandom localization of recombination events in human alpha satellite repeat unit variants: implications for higher-order structural characteristics within centromeric heterochromatin. *Mol. Cell. Biol.* **13**: 6520-6529.

Weber, B., Weissenbach, J. and Schempp, W. (1987). Conservation of human-derived pseudoautosomal sequences on the sex-chromosomes of the great apes. *Cytogenetics and Cell Genetics* 45: 26-29.

Weber, B., Weissenbach, J. and Schempp, W. (1988). X-Y crossing over in the chimpanzee. *Hum. Genet.* 80: 301-303.

Weber, J. L. (1990). Informativeness of human (dC-dA)n.(dG-dT)$_n$ polymorphisms. *Genomics* 7: 524-530.

Weber, J. L. and Wong, C. (1993). Mutation of human short tandem repeats. *Hum. Mol. Genet.* 2: 1123-1128.

Weir, B. S. (1990). *Genetic data analysis.* Sunderland, MA, Sinauer Associates, Inc.

Weisman-Shomer, P. and Fry, M. (1993). Quad, a protein from hepatocyte chromatin that binds selectively to guanine-rich quadruplex DNA. *J. Biol. Chem.* 268: 3306-3312.

Wellinger, R. J., Wolf, A. J. and Zakian, V. A. (1993). *Saccharomyces* telomeres acquire single-strand tg(1-3) tails late in s-phase. *Cell* 72: 51-60.

Wells, R. A., Germino, G. G., Krishna, S., Buckle, V. J. and Reeders, S. T. (1990). Telomere-related sequences at interstitial sites in the human genome. *Genomics* 8: 699-704.

White, M. A., Wierdl, M., Detloff, P. and Petes, T. D. (1991). DNA binding protein RAP1 stimulates meiotic recombination at the HIS4 locus in yeast. *Proc. Natl. Acad. Sci. USA* 88: 9755-9759.

Wilkie, A. O. M., Lamb, J., Harris, P. C., Finney, R. D. and Higgs, D. R. (1990). A truncated human chromosome-16 associated with alpha-thalassemia is stabilized by addition of telomeric repeat (TTAGGG)n. *Nature* 346: 868-871.

Wilkie, A. O. M., Higgs, D. R., Rack, K. A., Buckle, V. J., Spurr, N. K., Fischelghodsian, N., Ceccherini, I., Brown, W. R. A. and Harris, P. C. (1991). Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome-16. *Cell* 64: 595-606.

Wilkie, T. M. and Simon, M. I. (1991). Cloning multigene families with degenerate PCR primers. *Methods: A companion to methods in enzymology* 2: 32-41.

Willard, H. F. and Waye, J. S. (1987). Hierarchical order in chromosome-speicific human alpha satellite DNA. *Trends in Genetics* 3: 192-198.

Willard, H. F. (1990). Centromeres of mammalian chromosomes. *Trends in Genetics* 6: 410-416.

Williamson, J. R., Raghuraman, M. K. and Cech, T. R. (1989). Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell* 59: 871-880.

Winship, P. R. (1989). An improved method for directly sequencing PCR amplified material using dimethyl-sulphoxide. *Nucleic Acids Res.* 17: 1266.

Wolpoff, M. H. (1989). Multiregional evolution: the fossil alternative to Eden. *The human revolution: behavioural and biological perspectives on the origins of modern humans.* Edinburgh, Edinburgh University Press. 1: 62-108.

Wong, Z., Wilson, V., Jeffreys, A. J. and Thein, S. L. (1986). Cloning a selected fragment from a human DNA fingerprint: isolation of an extremely polymorphic minisatellite. *Nucleic Acids Res.* 14: 4605-4616.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A. J. (1987). Characterization of a panel of highly variable minisatellites cloned from human DNA. *Ann. Hum. Genet.* **51**: 269-288.

Wright, J. H., Gottschling, D. E. and Zakian, V. A. (1992). *Saccharomyces* telomeres assume a non-nucleosomal chromatin structure. *Genes & Development* **6**: 197-210.

Wright, J. H. and Zakian, V. A. (1995). Protein-DNA interactions in soluble telosomes from *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **23**: 1454-1460.

Wright, W. E. and Shay, J. W. (1992). Telomere positional effects and the regulation of cellular senescence. *Trends in Genetics* **8**: 193-197.

Yu, G. L., Bradley, J. D., Attardi, L. D. and Blackburn, E. H. (1990). *In vivo* alteration of telomere sequences and senescence caused by mutated *Tetrahymena* telomerase RNAs. *Nature* **344**: 126-132.

Yu, G. L. and Blackburn, E. H. (1991). Developmentally programmed healing of chromosomes by telomerase in *Tetrahymena*. *Cell* **67**: 823-832.

Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., Holman, K., mulley, J. C., Warren, S. T., Schlessinger, D., Sutherland, G. R. and Richards, R. I. (1991). Fragile-X genotype characterized by an unstable region of DNA. *Science* **252**: 1179-1181.

Yunis, J. J. and Prakash, O. (1982). The origin of man: a chromosomal pictorial legacy. *Science* **215**: 1525-1530.

Zahler, A. M., Williamson, J. R., Cech, T. R. and Prescott, D. M. (1991). Inhibition of telomerase by G-quartet DNA structures. *Nature* **350**: 718-720.

Zakian, V. A. and Pluta, A. F. (1989). Telomere formation in yeast. *Nature* **338**: 468-468.

Zangenberg, G., Huang, M. M., Arnheim, N. and Erlich, H. (1995). New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. *Nature Genet.* **10**: 407-414.

Zhong, Z., Shiue, L., Kaplan, S. and de Lange, T. (1992). A mammalian factor that binds telomeric TTAGGG repeats *in vitro*. *Mol. Cell. Biol.* **12**: 4834-4843.