The EMBO Journal vol.14 no.21 pp.5433–5443, 1995

Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human Xp/Yp telomere

D.M.Baird, A.J.Jeffreys and N.J.Royle[1]

Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

[1]Corresponding author

Sequences immediately adjacent to the human Xp/Yp telomere exhibit a high frequency of base substitutional polymorphisms, together with almost complete linkage disequilibrium, to create only a few diverged haplotypes. This sequence divergence has been used to develop a PCR-based system for mapping the distribution of the telomere (TTAGGG) and variant repeats (TGAGGG and TCAGGG) at the proximal end of the telomere repeat array. The distribution of these repeats is extremely variable. Almost all Xp/Yp telomeres are different, indicating a high mutation rate. Some telomere maps associated with the same flanking haplotype show similarities, identifying subsets of telomeres that share a recent common ancestry. Mechanisms underlying the rapid turnover of repeats at the proximal end of the Xp/Yp telomere include intra-allelic processes, such as slippage during replication. Inter-allelic exchanges may occur occasionally, but telomerase activity probably plays only a minor role in the germline turnover of proximally located telomere and variant repeats.

Keywords: hypervariable/polymorphism/pseudoautosomal/telomere/turnover

Introduction

The human Xp/Yp pseudoautosomal region (PAR1) is 2.6 Mb long, contains at least six genes, and is defined proximally by an Alu element and distally by a telomere (Rappold, 1993). In common with the proterminal regions of many human autosomes, PAR1 has a high GC content and contains many hypervariable GC-rich minisatellites. During male meiosis, chromosome pairing occurs between the homologous regions of the X and Y chromosomes, and PAR1 is the site of an obligatory recombination event (Cooke *et al.*, 1985; Simmler *et al.*, 1985; Schmitt *et al.*, 1994). Clearly, PAR1 serves a specialized function in the male germline but it also shares some of the physical and genetic properties that have been attributed to the proterminal regions of human autosomes (Royle *et al.*, 1988). Terminal restriction fragments, including the telomere of PAR1, have been cloned. A PAR1-specific minisatellite, which does not show any restriction fragment length variation, and a truncated copy of a short interspersed nuclear element (SINE) have been identified within 1 kb of the telomere (Figure 1A; Brown, 1989; Royle *et al.*, 1992).

Telomeres protect linear chromosomes from degradation and fusion to other chromosomes, and are thought to be a site of attachment to the nuclear matrix at times during the cell cycle (Biessmann and Mason, 1993). Telomeres of human chromosomes are composed of variable length arrays of TTAGGG repeat units, with the G-rich strand oriented 5′ to 3′ towards the telomere. Variant telomere repeat units such as TTGGGG and TGAGGG also occur, and tend to be located at the proximal ends of human telomeres (Allshire *et al.*, 1989). Telomere length decreases in human somatic tissues with each cycle of replication (Harley *et al.*, 1990; Hastie *et al.*, 1990), perhaps leading to the loss of functional telomeres and therefore to genome instability and cell death. As a result, telomere loss has been implicated in cellular senescence. The loss of telomere repeats can be counteracted by the activity of telomerase, which adds TTAGGG repeats *de novo* (Greider and Blackburn, 1989; Morin, 1989). Telomerase activity has been detected in the human germline, in many types of cancer, in immortal cell lines (Counter *et al.*, 1994; Kim *et al.*, 1994) but not in normal somatic cells. Therefore it has been suggested that activation of telomerase is an essential step in tumour progression and in the immortalization of cells in culture (Greider, 1994).

The internal structures of alleles at minisatellite loci have been analysed by mapping the distribution of repeat units which show sequence variation from the consensus repeat (Minisatellite Variant Repeat-PCR; Jeffreys *et al.*, 1991). This method has revealed the real extent of allelic variation at several minisatellites (Armour *et al.*, 1993; Monckton *et al.*, 1993; Neil and Jeffreys, 1993; Buard and Vergnaud, 1994) and has been used to characterize the processes involved in germline mutations at these loci (Jeffreys *et al.*, 1994).

To study allelic variation and sequence turnover within the PAR1 telomere, we characterized the flanking sequence polymorphisms and used these to develop a PCR-based method to amplify individual alleles of the pseudoautosomal telomere. The distribution of telomere TTAGGG and variant TGAGGG and TCAGGG repeats at the proximal end of this telomere, determined by telomere variant repeat (TVR) mapping by PCR, has revealed extensive allelic variation. It should be possible to exploit this method to study the mutation processes that underlie the allelic variation and molecular changes that occur at the proximal end of the Xp/Yp telomere, including those operating during cellular senescence in somatic tissues.

Results

Identification of polymorphisms in the DNA adjacent to the Xp/Yp telomere

Single-stranded conformational polymorphism analysis (Orita *et al.*, 1989) of the 480 bp proximal to the start of

the human Xp/Yp telomere revealed a high level of variation (data not shown). To determine the nature of this variation, a sequence analysis of PCR products (primers TSK8C to TSK8G and TSK8E to TSK8B; Figure 1A) covering 850 bp of DNA flanking the telomere was undertaken. Among the 32 Caucasian and 21 African DNAs analysed, 13 and 17 polymorphic sites were identified, respectively. Therefore the frequency of base substitutional polymorphisms is one per 65 bp in Caucasians and one per 50 bp in Africans. In addition, a 10 bp insertion/deletion polymorphism was identified in African DNAs (Table I and Figure 1B) plus two variant positions (Figure 1B, –217 and –826) which occurred only once among the DNAs sequenced. The 18 polymorphic positions identified in the two populations are distributed throughout the 850 bp sequenced (Figure 1 and Table I). The polymorphisms include 10 transitions and eight transversions; six of the polymorphisms, including the –826 variant position, have arisen at CpG doublets presumably by mutation of the C residue.

Strong linkage disequilibrium adjacent to the Xp/Yp telomere

Some of the polymorphisms create or destroy restriction enzyme sites. Four sites (–415, –427, –652 and –842) located within and proximal to the minisatellite were selected for further analysis by the PCR amplification of genomic DNA using primers TSK8C + TSK8G (Figure 1A) followed by digestion with the appropriate enzymes (*Mbo*II, *Taq*I, *Ava*II or *Dde*I, respectively). Surprisingly, a number of unrelated individuals were homozygous and others heterozygous at all four polymorphic positions. Therefore a haplotype analysis was carried out with an amplification refractory mutation system (ARMS) PCR primer (TS-842C) for the specific amplification of one of the two alleles (Newton *et al.*, 1989) at the –842 polymorphic position. This was followed by a restriction fragment length polymorphism (RFLP) analysis to assay the –415, –427 and –652 positions. A haplotype analysis of the four polymorphic sites revealed only three (A, B and C) of the 16 possible haplotypes (Tables I and II) and haplotypes B and C differ only at position –427. There is almost complete linkage disequilibrium across the four polymorphic sites. The haplotypes vary in frequency in different populations (Table II) and are in Hardy–Weinberg equilibrium in each population tested.

This analysis was extended to the –176 polymorphism, which creates a *Dde*I RFLP within the truncated SINE (see Figure 1). Among the 79 Caucasian DNAs tested, all DNAs heterozygous for the haplotypes (A, B or C) were also heterozygous at the –176 polymorphism; DNAs homozygous for the A or B haplotypes were homozygous at –176, with either a G or T nucleotide respectively (Table I). In addition, ARMS assays were established for the –13 and –30 polymorphic positions. Amplification from the –13 position using the allele-specific primer TS–13AR to primer TSK8C was carried out; the products were digested with *Dde*I to assay the –176 and –842 polymorphisms. Again, the –176 polymorphism was in linkage disequilibrium with haplotypes A, B and C in all

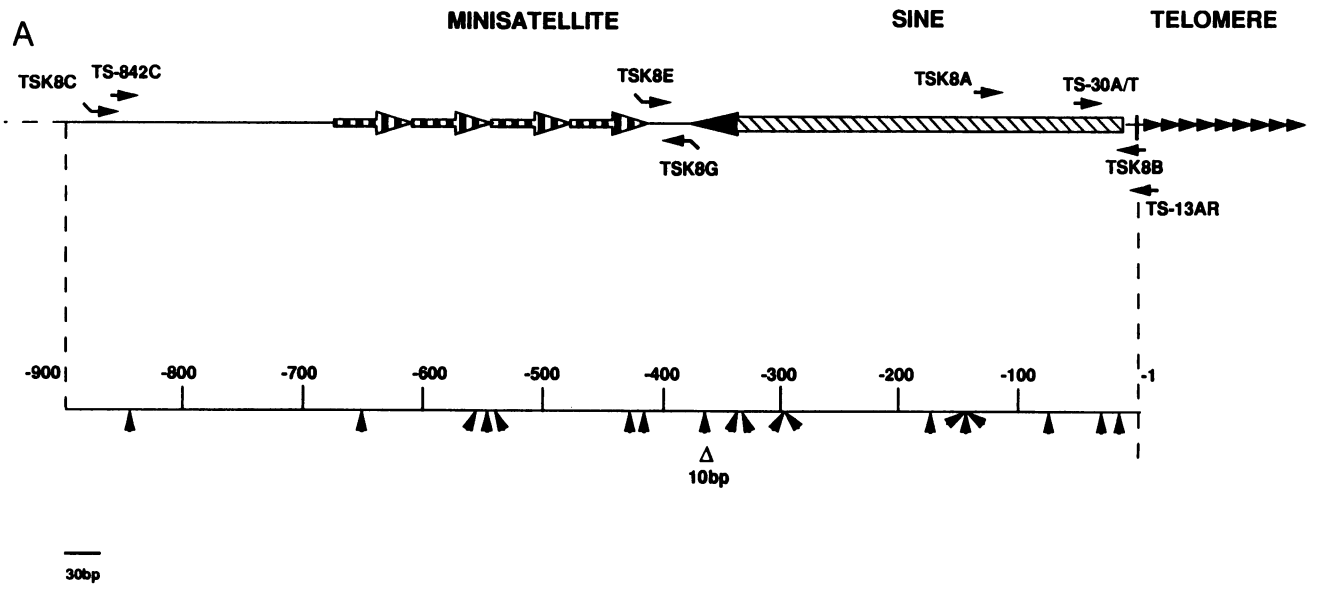


Fig. 1. Telomere junction of the Xp/Yp pseudoautosomal region. (**A**) The distribution of polymorphic positions (solid arrowheads on lower line) in the DNA adjacent to the Xp/Yp telomere repeat array (series of solid arrowheads pointing right), which contains a SINE (hatched rectangle) and a minisatellite (striped arrows). The locations of the PCR primers used in the analysis are shown: (solid horizontal arrow) universal and allele-specific primers; (solid horizontal arrow with a bent tail) primers including a noncomplementary 5′ tail. The sequence is numbered from the first base adjacent to the telomere (–1). (**B**) Sequence of the DNA adjacent to the Xp/Yp telomere numbered from the first base (–1) adjacent to the first repeat of the telomere. The sequence of the minisatellite is shown as capital letters and the SINE in italics. The polymorphic positions are numbered above the sequence (–13 to –842) and alternative bases (bold type) at the polymorphic sites are shown. The sequence variation at –217 was observed only once in the Caucasian DNAs, and that at –826 only once in the African DNAs sequenced; they are not shown in (A). The restriction enzymes for the sites which create an RFLP are shown, and the 10 bp deleted in some African DNAs is underlined. The positions of the PCR primers are shown above (identical) or below (complementary) the sequence. An asterisk at the end of a primer represents a noncomplementary 5′ primer tail (see Materials and methods).

79 Caucasians tested. The –13 and –30 polymorphisms also showed concordance with the flanking haplotypes, such that all 78 Caucasian alleles of haplotype B or C had nucleotide A at the –13 and –30 positions, and all except three of the 80 haplotype A alleles had nucleotide T at these two positions. The three alleles designated haplotype A that revealed additional heterogeneity only differed across the –30 and –13 polymorphisms: one variant haplotype A allele had nucleotide A at the –13 position instead of T; the two others had nucleotide A instead of T at both the –30 and –13 positions.

Almost complete linkage disequilibrium has therefore been identified across seven polymorphic sites (–13, –30, –176, –415, –427, –652 and –842) adjacent to the Xp/Yp telomere in Caucasian DNAs. Similarly, nearly complete linkage disequilibrium has been demonstrated across four polymorphic sites (–415, –427, –652 and –842) in Africans, Japanese and Afro-Caribbeans, with only three common haplotypes in each population (Table II).

Determining the DNA sequence divergence between haplotypes

To determine whether the intervening polymorphic sites were also in strong linkage disequilibrium, three Caucasian DNAs homozygous for haplotype A and three homozygous for haplotype B were sequenced. The intervening polymorphic positions were all homozygous (Table I) and in complete linkage disequilibrium with the haplotypes (A and B). In addition, three haplotype C alleles were sequenced after allele-specific amplification from individuals heterozygous for haplotypes A and C; the sequences of the three Caucasian haplotype C alleles were again identical (Table I). The 13 polymorphisms identified in the 850 bp adjacent to the Xp/Yp telomere are in almost complete linkage disequilibrium in Caucasian DNAs, and

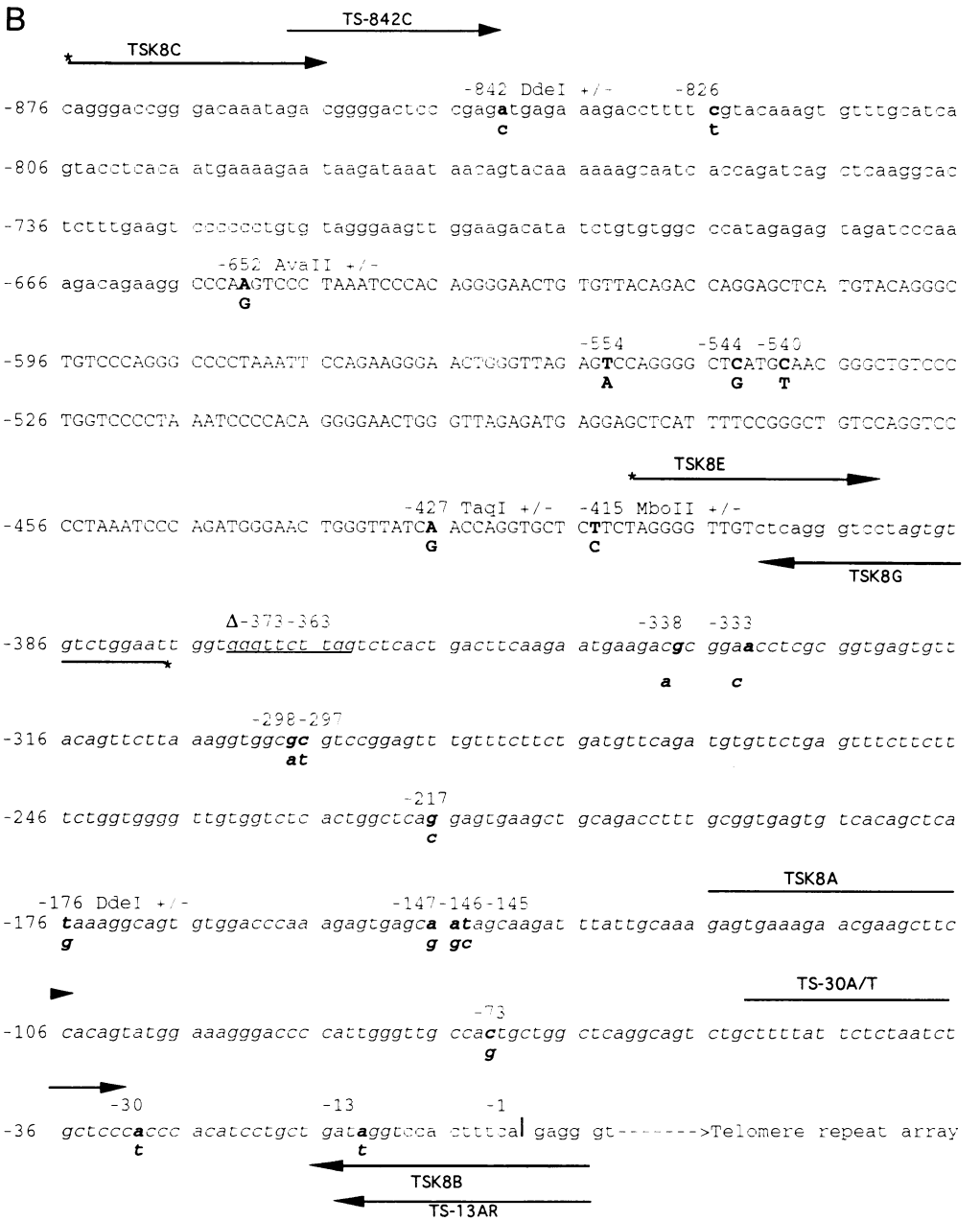


Table I. Polymorphisms in the DNA adjacent to the Xp/Yp telomere

Position from first telomere repeat	−842 *Dde*I +/−	−652 *Ava*II +/−	−554	−544	−540	−427 *Taq*I +/−	−415 *Mbo*II +/−	Δ373–363[a]	−338	−333	−298	−297	−176 *Dde*I +/−	−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**
African haplotypes																			
A	**C**	**G**	A	C	C	**G**	**C**	−	A	C	A	T	G	G	G	C	G	nt	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	T	**G**	**T**	+	G	A	G	C	T	A	G	T	C	nt	nt

The polymorphic positions are numbered from the telomere (right of the table). The assays for the polymorphic sites analysed extensively are shown below. The top two rows of the table show the bases that occur at the polymorphic positions (5′→3′→telomere) in Caucasian and African DNAs. Full haplotypes across all the polymorphisms are shown at the bottom of the table. These were obtained by sequence analyses of six Caucasian haplotype A alleles, six Caucasian haplotype B alleles and three Caucasian haplotype C alleles; six African haplotype A alleles, six African haplotype C alleles and one African haplotype B allele; and six Afro-Caribbean haplotype B alleles (seven haplotype B alleles in total). The first column shows the haplotypes (A, B or C) assigned by an analysis of the −415, −427, −652 and −842 positions. nt, not tested.
[a]The 10 bases between −363 and −373 are present (+) or absent (−), as indicated.

Table II. Frequencies of haplotypes at the Xp/Yp pseudoautosomal telomere-junction and at the −1888 *Nla*IV polymorphism

Haplotype	−842 *Dde*I	−652 *Ava*II	−427 *Taq*I	−415 *Mbo*II	Caucasians (n = 78)	Japanese (n = 105)	Africans (n = 28)	Afro-Caribbeans (n = 37)
A	C	G	G	C	0.513	0.286	0.750	0.730
B	A	A	A	T	0.442	0.710	0.018	0.149
C	A	A	G	T	0.045	0.005	0.232	0.122
−1888 *Nla*IV								
+					0.571	0.223	0.352	0.419
−					0.429	0.777	0.648	0.581

Haplotypes A, B and C were determined by an analysis of four polymorphic positions (−842, −652, −427 and −415). The frequency of the three haplotypes and the −1888 polymorphisms is shown for each population; each polymorphism is in Hardy–Weinberg equilibrium in the populations tested (data not shown); n is the number of individuals tested for each population.

the sequence divergence between the A and B haplotypes is therefore 1.5%.

Three African DNAs homozygous for haplotype A and three homozygous for haplotype C were sequenced. Seven haplotype B alleles, one from an African DNA heterozygous for haplotypes A and B and six from three Afro-Caribbean DNAs homozygous for haplotype B, were also sequenced (Table I). The sequence at the intervening polymorphisms was the same within a group of sequenced African alleles designated haplotype A, B or C, except at the −544 and −540 positions. One African DNA homozygous for haplotype A and one homozygous for haplotype C were both heterozygous at the −544 and −540 positions but not at any of the other polymorphic sites. Further sequence analyses (not included in Table I) have identified another African DNA homozygous for haplotype A, and homozygous at all the intervening polymorphisms, but with nucleotides G and T instead of C and C at the −544 and −540 positions, respectively. Apparently, a switch across the −544 and −540 polymorphisms in the minisatellite adjacent to the Xp/Yp telomere has occurred between the African A and C haplotypes (Figure 1 and Table I). The switch could have arisen as an inter-allelic conversion-like exchange between the minisatellites in African haplotypes A and C. The identification of two additional haplotypes in a limited number of African DNAs suggests that there is greater heterogeneity within this population than in Caucasians. Nevertheless, there is strong linkage disequilibrium across the 16 base substitutional polymorphisms and one 10 bp insertion/deletion polymorphism in the 820 bp (not including the −30 and −13 positions) of telomere-adjacent DNA in Africans. Therefore, sequence divergence up to 2.1% exists between the African A and B haplotypes.

The sequence analysis of a limited number of alleles has shown that B haplotypes are identical in Africans/Afro-Caribbeans and Caucasians. The African and Caucasian C haplotypes are also the same, although the sequenced African C haplotypes show some heterogeneity because of an apparent exchange between the African A and C haplotypes across the −544 and −540 positions that was not observed in Caucasians. In contrast, the African and Caucasian A haplotypes (Table I) differ at eight of the 18 sites that are known to be polymorphic. Two other variant

Table III. Linkage disequilibrium between the flanking haplotypes of the Xp/Yp telomere and the –1888 polymorphic position in Caucasians

–842, –652, –427, –415 haplotype	–1888 *Nla*IV +/–	Observed frequency (n = 156)	Observed No.	Expected No.	D value	r value
A	+	0.462	72	46	0.169[b]	0.683
A	–	0.051	8	34	–0.169[b]	–0.683
B	+	0.109	17	39	–0.143[b]	–0.582
B	–	0.333	52	30	0.143[b]	0.582
C	+	0.000	0	4	–0.026[a]	–0.253
C	–	0.045	7	3	0.026[a]	0.253

The disequilibrium coefficients, D, for all allele combinations are shown. $D_{uv} = p_{uv} - p_u q_v$, where p_{uv} is the frequency of haplotype A_uB_v, p_v is the frequency of allele A_v at the first locus and q_v is the frequency of allele B_v at the second locus. D can vary between –0.25 and +0.25. The D values have been tested for significant deviation from random association using χ^2 statistics: [a]$P < 0.05$, [b]$P < 0.001$. In an overall test for significant deviation of the D values from zero (Weir, 1990), total $\chi^2 = 37.24$ and with two degrees of freedom $P < 0.01$. The linkage disequilibrium coefficients have also been converted to r values after standardization for gene frequencies; r can vary between –1 and +1.

positions were detected during the sequence analysis: one at position –826 (C → T) occurred in an African A haplotype chromosome and the other at position –217 (G → C) occurred in a Caucasian A haplotype chromosome (Figure 1B). These additional variants at –826 and –217 may have arisen recently and could be limited to African and Caucasian populations, respectively.

Defining the proximal limit of the high frequency of base substitutional polymorphisms

Sequence information from clone T7A1/4 (Brown *et al.*, 1990), which includes 2 kb of DNA adjacent to the Xp/Yp telomere, was used to design primers ~2.0 and 1.5 kb from the telomere. The direct sequence analysis of 320 bp between these primers (TSK8J and TSK8K) from 17 Caucasian and 11 African DNAs revealed one additional polymorphic position (–1888, G or A) common to both populations and another polymorphic position (–1897, G or A) only in the African DNAs. The frequency of base substitutional polymorphisms in this region is reduced compared with the 850 bp adjacent to the telomere in Caucasians and Africans, but the difference is not significant (for Caucasians $P = 0.224$ and for Africans $P = 0.176$ using Fisher's two-tailed, exact test). The –1888 polymorphism can be detected by the presence or absence of an *Nla*IV restriction site, and was found to be in Hardy–Weinberg equilibrium in Caucasian, Japanese, African and Afro-Caribbean populations (Table II). A haplotype analysis between the –1888 polymorphism and flanking haplotypes A, B and C was undertaken in Caucasians (Table III); strong but incomplete linkage disequilibrium (Weir, 1990) between the flanking haplotypes and the –1888 polymorphism was identified in this population. A comparison of the –1888 allele with the flanking haplotype frequencies in Africans and Afro-Caribbeans (Table II) suggests that there is much weaker linkage disequilibrium between the flanking haplotype and –1888 polymorphism in these populations. It seems likely that the very high frequency of base substitutional polymorphisms combined with near complete linkage disequilibrium is confined to the extreme end of the PAR1 adjacent to the telomere.

TVR mapping by PCR

These base substitutional polymorphisms have enabled us to develop an assay for the distribution of TTAGGG and variant repeats in the proximal 1 kb of the Xp/Yp telomere itself. Briefly, the method includes the amplification of genomic DNA using a radioactively labelled allele-specific flanking primer, together with one of three 'tagged' telomere or variant repeat primers. Amplification from the –30 polymorphism, within the SINE adjacent to the Xp/Yp telomere, by one of the allele-specific primers (TS–30A or TS–30T) generates single allele telomere maps in individuals heterozygous at the –30 position. The amplified products are resolved on a denaturing polyacrylamide gel and detected by autoradiography. Ladders of bands based on a 6 bp repeat unit are produced and the interspersion pattern of the TTAGGG, TGAGGG and TCAGGG repeats can be determined (Figure 2).

The TVR primers TAG-TELW, TAG-TELX and TAG-TELY are all composed of four repeat units; the 3′ base mismatches the second position of a 6 bp repeat unit such that TAG-TELW, TAG-TELX and TAG-TELY anneal to TTAGGG, TGAGGG and TCAGGG repeat types, respectively. The corresponding position in the penultimate repeat was made ambiguous (N) to aid equal amplification from different repeat types. The 5′ repeats of each TAG-TEL primer should anneal to TTAGGG, TGAGGG or TCAGGG repeats with similar but reduced efficiency because two repeats complementary to TAAGGG have been included. Our investigations have indicated that TAAGGG repeat types occur infrequently in human telomere repeat arrays; therefore they are unlikely to contribute to the preferential amplification of some repeats within a telomere repeat array by the TAG-TEL primers. In addition, a 20 nucleotide noncomplementary tail has been added to the 5′ end of each TVR primer (Jeffreys *et al.*, 1991), but the TVR–PCRs do not include the addition of a tail primer to drive the amplification and prevent collapse of the tandemly repeated products. However, the presence of a noncomplementary tail was found to promote the equal amplification of products from each repeat unit within the telomere repeat array and reduce the generation of additional bands at the end of a block of uniform repeat types.

Figure 3 shows the allele-specific amplification of telomeres from unrelated individuals heterozygous at the –30 flanking polymorphic position. TVR–PCR products are resolved as a 6 bp ladder, with a band appearing in one of the three tracks (T, G or C) for each repeat position. Some repeats fail to amplify, presumably because of the presence of additional telomere repeat sequence variation; such unknown repeats have been termed N-type repeats. About 120 repeat units can be resolved on denaturing

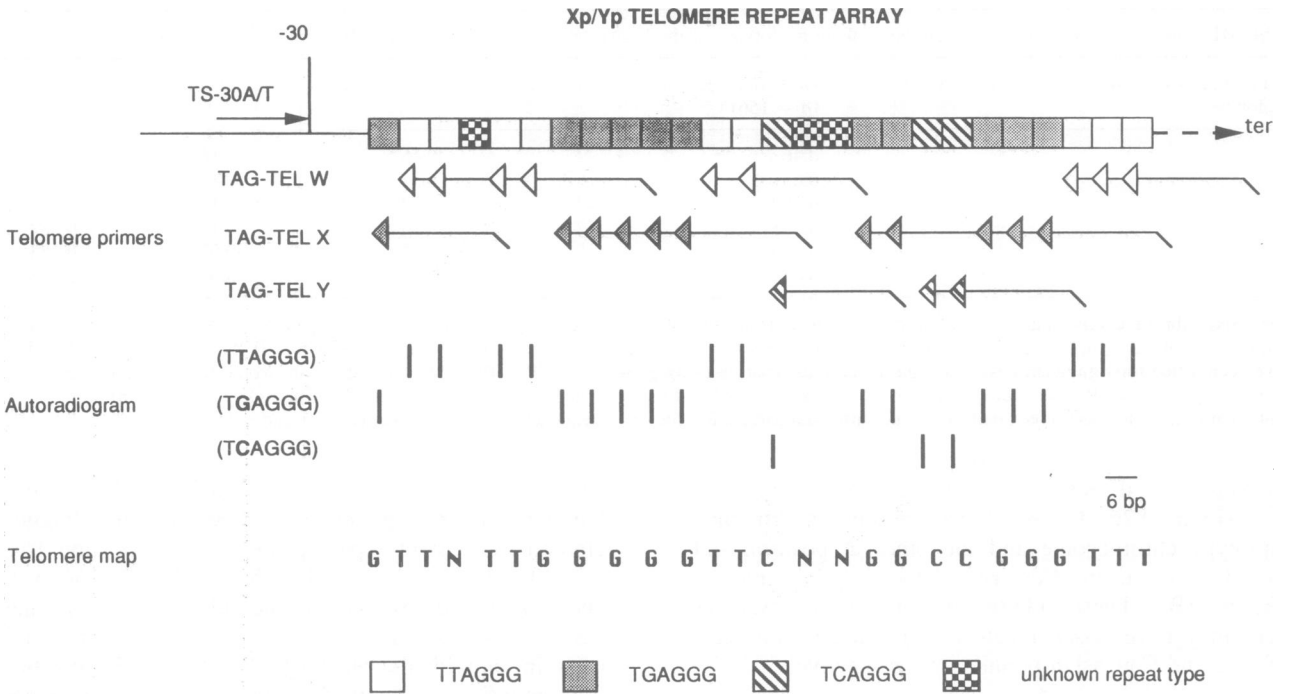


Fig. 2. The principle of telomere map generation. Allele-specific amplification is achieved in individuals heterozygous for the flanking haplotypes (A and B or A and C) from the –30 polymorphic position using a radioactively labelled primer TS–30T or TS–30A for the amplification of haplotype A or haplotypes B and C alleles, respectively. Amplification from the telomere TTAGGG (white square) or variant TGAGGG (grey square) and TCAGGG (hatched square) repeat types is achieved in separate reactions with primers TAG-TELW (bent white arrow), TAG-TELX (bent grey arrow) or TAG-TELY (bent hatched arrow). The products are resolved on a denaturing polyacrylamide gel into ladders of bands based on a 6 bp repeat unit length. The three tracks are read into a telomere code of T (TTAGGG), G (TGAGGG), C (TCAGGG) and N (unknown) repeat types. (Chequered square) N-type repeat.

polyacrylamide gels and the pattern of bands can be readily converted to telomere codes of T-, G-, C- and N-type repeats for database analysis (Figures 2 and 3).

Mendelian inheritance of Xp/Yp telomeres

The Mendelian inheritance of Xp/Yp telomere maps has been verified in families, and the segregation of the paternal telomere maps in Centre d'Etude du Polymorphisme Humain (CEPH) family 1423 is shown in Figure 4. In this family, the grandfather (1423–11) was heterozygous (A/T) at the –30 position and a telomere map (designated A′) was generated by amplification with the allele-specific primer, TS–30A. The grandmother (1423–12) was homozygous (A/A) at the –30 position and generated a diploid pattern from the two superimposed telomere maps, A and A″. The A″ telomere contributed only a few bands to the diploid map, probably because many N-type repeats were present at the proximal end of this telomere. The father (1423–01) was also homozygous at the –30 position and inherited the paternal A′ and maternal A telomere maps, which produced a diploid pattern different from that seen in his mother. The haploid telomere map amplified in the son (1423–03) was identical to the A′ telomere map in the grandfather, while the haploid telomere map (A) in the daughter can be seen as part of the diploid telomere pattern in her father and grandmother. The telomere maps therefore segregate as Mendelian markers both in this family and in six other CEPH families (data not shown). The extreme variability at the proximal end of the PAR1 telomere must be maintained by a relatively high *de novo* mutation rate, but no mutant alleles have been identified to date among the seven families examined.

Telomere variability

Telomere maps were established for 65 alleles from the parents of the CEPH family DNAs heterozygous for the –30 flanking polymorphism (Figure 5). One of the telomere maps associated with flanking haplotype A and nine associated with haplotype B were composed almost entirely of TTAGGG repeats. Five of the haplotype B alleles were indistinguishable but all other telomeres were different, giving 61 different allelic structures in 65 telomeres mapped. All haplotype A alleles began with at least two N-type repeats, and therefore included 12 bp of unknown sequence. In contrast, all telomeres associated with flanking haplotype B and C began with a G-type repeat, with the exception of haplotype B telomere 1329102 which like most of the A haplotype telomeres began with two N-type repeats. The variant repeats within the telomeres showed a strong tendency to cluster, to produce, for example, runs of G- and N-type repeats in haplotype A alleles and C-type repeats in haplotype B alleles. In some cases there was evidence of a higher order repeat structure, e.g. runs of a TGG three-repeat motif in some haplotype C alleles (as seen in 4501). TVR–PCR with an additional TVR primer has shown that many N-type repeats in haplotype A alleles are TTGGGG (data not shown). Occasionally a change in the 6 bp periodicity of the ladder of bands was detected, e.g. in haplotype A 1201 and related telomeres (Figure 5), and sequence analysis has shown that additional variants exist amongst

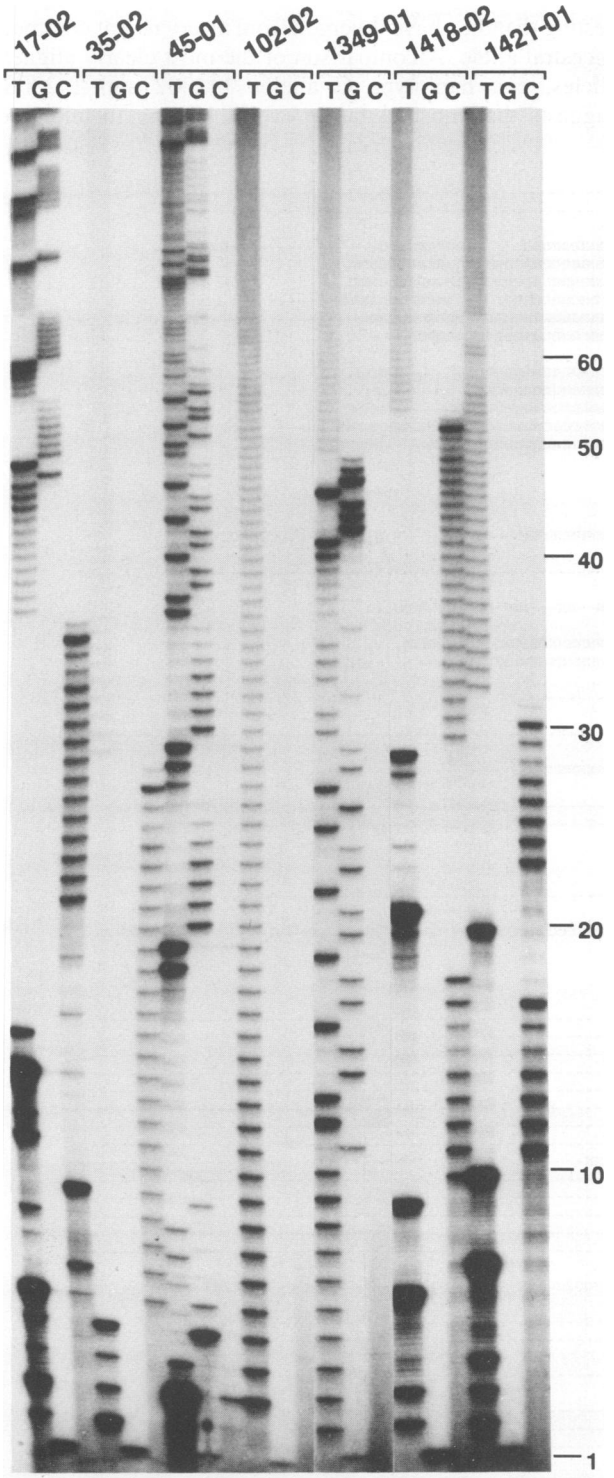


Fig. 3. Examples of telomere maps by TVR–PCR. Each set of three tracks, T, G and C, represents a single telomere map from the seven unrelated CEPH individuals indicated. The telomeres have been amplified with the TS–30A primer and are therefore associated with B or C flanking haplotypes. A band is present at ~6 bp intervals in one or none of the tracks. The distribution of T (TTAGGG), G (TGAGGG), C (TCAGGG) and N (unknown) repeats can be read from the autoradiogram. The first repeat of the telomeres shown is amplified by TAG-TELX to give a product of 99 bp in the G tracks. The scale shows the number of telomere repeats.

other N-type repeats. Clearly, the underlying sequence variation at the proximal ends of the Xp/Yp telomeres is greater than is currently detected by TVR–PCR.

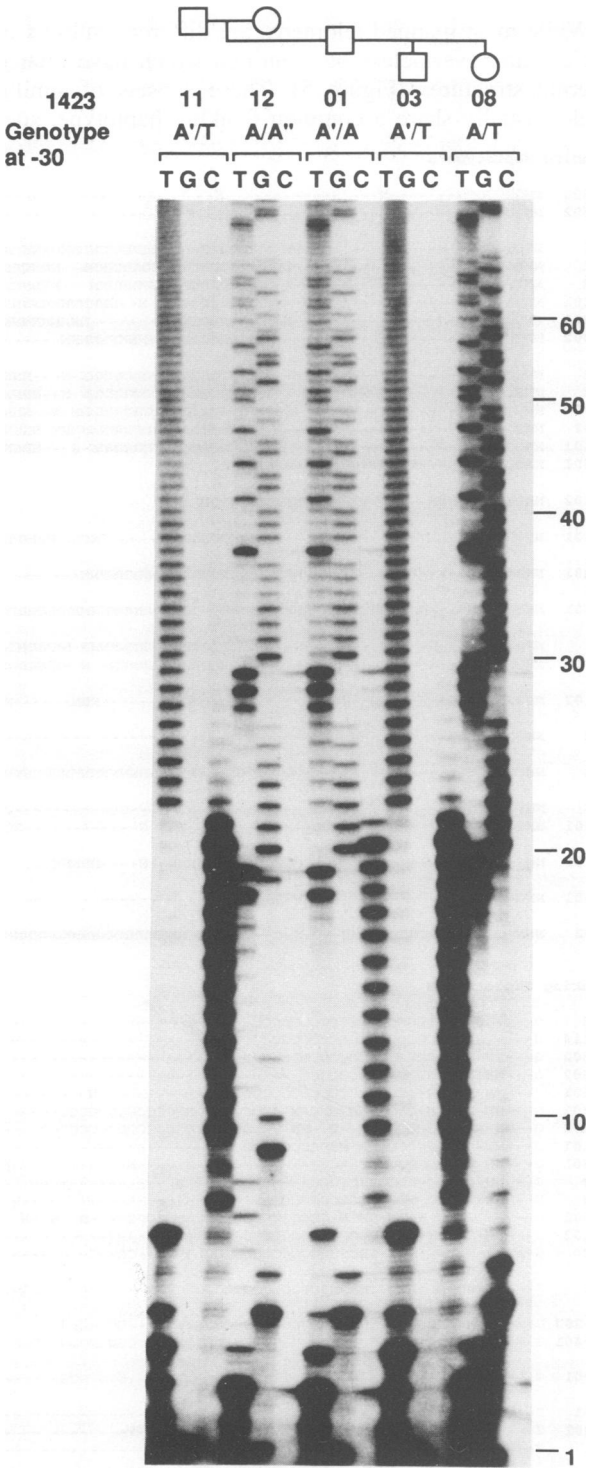
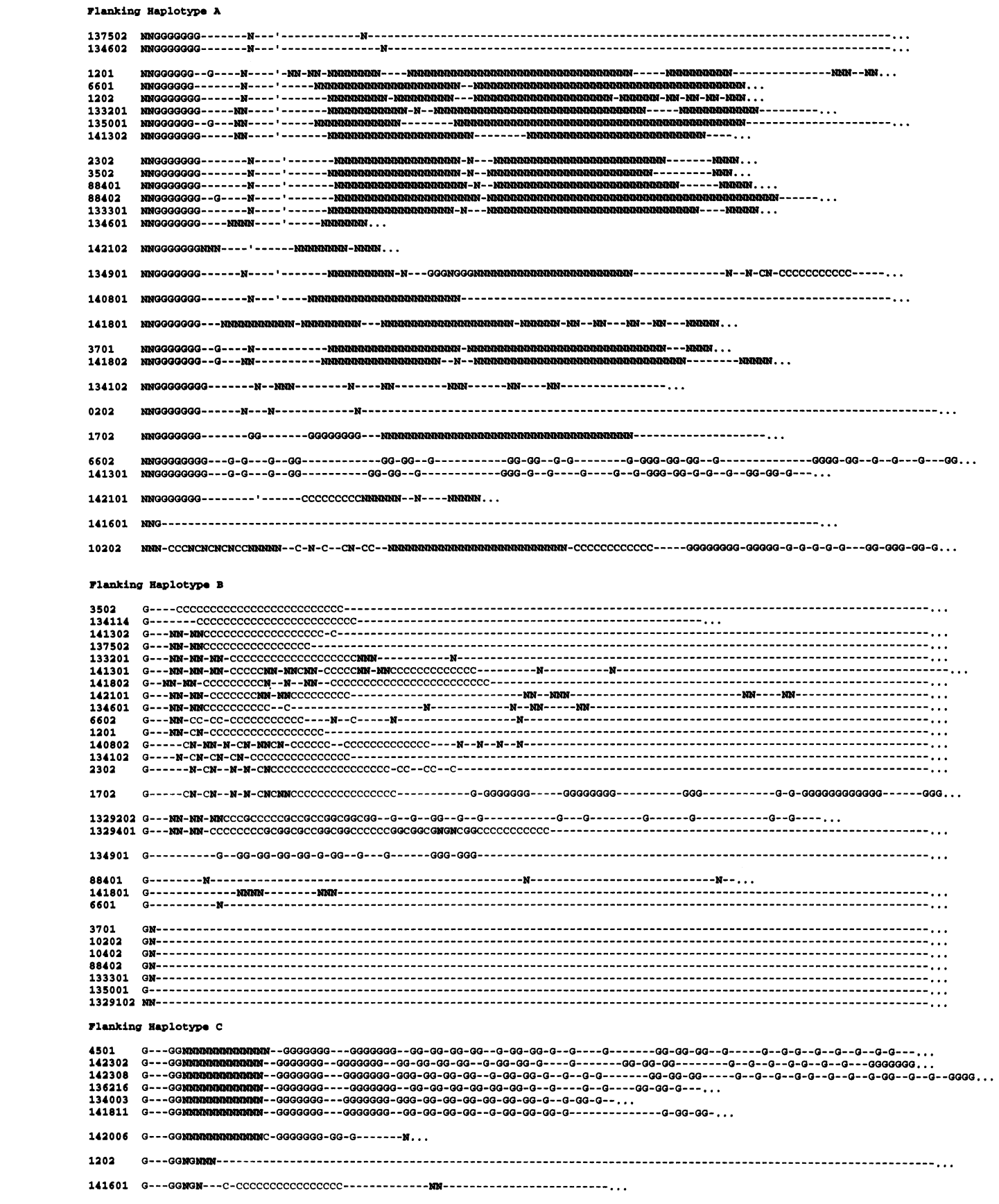


Fig. 4. Segregation of telomeres within a family. Telomere maps were generated using the TS–30A flanking primer from the grandfather (FF = 11), grandmother (FM = 12), father (F = 01), son (S = 03) and daughter (D = 08) from the CEPH 1423 family. The grandfather was heterozygous (AT) at the –30 position and a haploid telomere map (A′) was produced. The grandmother was homozygous (AA) at the –30 position and therefore a diploid telomere pattern was produced from the combined maps of telomeres designated A and A″. The father was also homozygous at the –30 position and produced a diploid telomere pattern, from the A′ and A telomeres in his parents. The two children were both heterozygous at the –30 position: the son has inherited the A′ telomere from his father while the daughter inherited the A telomere map. Each telomere map is deduced from three tracks (T, G and C) for the distribution of TTAGGG, TGAGGG and TCAGGG repeat types, respectively.

While most mapped telomeres are different, subsets of alleles can nevertheless be identified which have related internal structures (Figure 5). These subsets of similar alleles usually share a common flanking haplotype, suggesting that they have diverged from a very recent common ancestral allele. A comparison of the most clearly aligned alleles, e.g. haplotype C alleles 142302 and 142308, suggests that most mutation events leading to telomere



diversity involve the localized gain or loss of one or a few repeat units to create blocks of uniform repeat types. If true, this pattern of mutation is most likely supported by local replication slippage events involving one or a few repeat types.

Discussion

The frequency of base substitutional polymorphisms in noncoding regions of the human nuclear genome is about one in 235 bases (Nickerson *et al.*, 1992), although in a few regions, such as the DQA locus in the HLA gene complex (Gyllensten and Erlich, 1988), the frequency of polymorphic sites is much higher. In the 850 bp adjacent to the Xp/Yp telomere we have identified a higher than average frequency of base substitutional polymorphisms (one in 65 bp in Caucasians; one in 50 bp in Africans). However, there is almost complete linkage disequilibrium across polymorphic sites in the 850 bp, and therefore only a small number of haplotypes within a population. An analysis of other clusters of biallelic polymorphisms in noncoding regions of the human genome has revealed multiple haplotypes, with only partial linkage disequilibrium over short stretches of DNA sequence (Nickerson *et al.*, 1992; Monckton *et al.*, 1993). Partial linkage disequilibrium has been identified between the Y chromosome and an allele at the pseudoautosomal minisatellite, DXYS17 (Decorte *et al.*, 1994), and strong linkage disequilibrium between Y chromosome-specific sequences and a polymorphism a few hundred bases into the pseudoautosomal region (Ellis *et al.*, 1990), but the gradient of recombination across the pseudoautosomal region in the male germline may explain these associations.

Linkage disequilibrium can arise by admixture, genetic drift, selection and suppression of recombination, but it is not clear which of these effects have influenced the establishment of such strong linkage disequilibrium adjacent to the Xp/Yp telomere. There has been a high turnover of sequences at the ends of higher primate chromosomes during evolution (Royle *et al.*, 1994), and our preliminary investigations suggest that the sequence organization at the end of Xp/Yp and the location of the telomere are different between chimpanzees, gorillas and humans. The linkage disequilibrium and high frequency of base substitution polymorphisms adjacent to the human Xp/Yp telomere have resulted in haplotypes that differ by as much as 2.1%; this suggests that these haplotypes are ancient and raises questions as to how these ancient haplotypes have been preserved in modern human populations and why intermediate haplotypes appear to be so rare. Further investigation of these sequences and the location of the Xp/Yp telomere in chimpanzees and gorillas may help to resolve these questions.

The close proximity of some polymorphic sites to the start of the Xp/Yp telomere repeat array has facilitated the development of TVR mapping by PCR. Remarkably, allele-specific amplification of Xp/Yp telomeres can be achieved by TVR–PCR when the repeat primers are able to anneal at all telomeres and the polymorphic site used for allele-specific amplification lies within a SINE. There are an estimated 20 000 copies of this SINE family in the genome (La Mantia *et al.*, 1989), but sequence divergence between copies may have reduced the potential number of annealing sites for the TS–30A and TS–30T primers within the genome. In addition, the likely absence of other copies of this SINE family close to other telomeres could explain the specific amplification from the Xp/Yp telomere that we have observed.

TVR–PCR is a novel system for exploring allelic diversity in telomeres and mechanisms that underlie the evolution of sequences at the proximal end of a telomere. The extreme allelic variation revealed at the proximal end of Xp/Yp telomeres in Caucasians suggests a high underlying mutation rate, although *de novo* mutants have yet to be identified. An analysis of mouse DNA has revealed telomere length hypervariability within and between inbred strains of mice. In addition, mouse pedigree analysis has shown that the length variation in the telomeres and associated restriction fragments is supported by a high mutation rate (Kipling and Cooke, 1990; Starling *et al.*, 1990). The majority of human Xp/Yp telomeres we have mapped indicate that in the germline at least, the proximal ends of telomeres are unperturbed by the activity of telomerase which adds (TTAGGG) repeats onto the terminus of telomeres. A comparison of closely related alleles suggests that the proximal end of a telomere most likely evolves by slippage, plus the occasional base substitution that introduces a novel variant repeat into the array. A strong association between related telomere maps and flanking haplotypes also suggests that telomeres are evolving largely along haploid lineages, consistent with an intra-allelic process such as slippage. However, there is some indication that exchanges between different alleles have occurred occasionally. For example, the telomere map of 134901 associated with haplotype B is similar to other haplotype B-associated telomeres, but after 11 repeats the map becomes more similar to telomere maps such as 4501 found in association with flanking haplotype C. It is not known whether these infrequent exchanges arise through inter-allelic recombination or through a conversion-like process known to generate diversity at minisatellites (Armour *et al.*, 1993; Monckton *et al.*, 1993; Neil and Jeffreys, 1993; Jeffreys *et al.*, 1994). Whatever the basis of such exchanges, they cannot often involve recombination within the DNA immediately adjacent to this telomere because such exchanges would disrupt the existing strong linkage disequilibrium.

TVR–PCR is a DNA typing system that may be useful in forensic DNA analysis, particularly for degraded DNA when the interspersion patterns of the short repeat units

Fig. 5. Hypervariability of proximal Xp/Yp telomere arrays. Single allele telomere maps, obtained from the indicated CEPH individuals heterozygous at the –30 flanking polymorphism, were coded for the distribution of T (TTAGGG), G (TGAGGG), C (TCAGGG) and N (unknown) repeat types. The G-, C- and N-type repeats are shown in red, green and black, respectively. To aid visualization of the different interspersion patterns, the T-type repeats are shown as blue bars. The telomere maps obtained from 65 telomeres have been grouped according to their flanking haplotype. Haplotype A alleles have nucleotide T at the –30 position and were amplified by the TS–30T primer, whereas haplotypes B and C have nucleotide A at the –30 position and were amplified by the TS–30A primer. All except one of the haplotype B and C telomeres begin with a G-type repeat. The unknown sequence at the beginning of the haplotype A telomeres is represented by two N-type repeats. A prime (′) has been introduced to indicate a minor change in the 6 bp periodicity of some telomere maps.

could still be determined. TVR–PCR may also be suitable for automation using fluorescently labelled flanking primers and typing on an automated sequencer. The technique could also be applied to studies of somatic turnover of telomeres in, for example, senescing cell lines that do not express telomerase, and it could, in principle, be developed for other human telomeres and for other species whose telomeres are composed of TTAGGG or similar repeats.

Materials and methods

Genomic DNAs

Caucasian DNAs were lymphoblastoid DNAs from 80 unrelated individuals which constitute the parents of the CEPH families. African and Afro-Caribbean DNAs were extracted from whole blood samples which had been collected in the UK. The Japanese DNA samples from whole blood were a gift from K.Tamaki (Jeffreys *et al.*, 1991).

Primers and PCR

Primers were synthesized with a 308B DNA synthesizer (Applied Biosystems) using reagents from Cruachem. Primer sequences are shown below; the noncomplementary 5′ tails of primers TSK8C, TSK8E, TSK8G, TAG-TELW, TAG-TELX and TAG-TELY are shown in bold and the restriction sites in the tails are italicized. PCRs were performed using the buffer described previously (Jeffreys *et al.*, 1991) with 1 μM of each primer and 0.1 U/μl *Taq* polymerase in a Perkin Elmer Cetus 9600 thermal cycler, unless otherwise stated.

Primers for PCR

TSK8A, 5′-GAGTGAAAGAACGAAGCTTCC-3′; TSK8B, 5′-CCCTCTGAAAGTGGACCTAT-3′; TSK8C, 5′-**GC*GGTACC***AGGGACCGGGACAAATAGAC-3′ (restriction enzyme *Kpn*I); TSK8E, 5′-**GC*GGTACC***TAGGGGTTGTCTCAGGGTCC-3′ (restriction enzyme *Kpn*I); TSK8G, 5′-**CG*GAATTC***CAGACACACTAGGACCCTGA-3′ (restriction enzyme *Eco*RI); TSK8J, 5′-GAATTCCTGGGGACTGCGGATG-3′ (–1986); TSK8K, 5′-CATCCCTGAAGAAGCATCTTGGCC-3′ (–1566).

Allele-specific primers

TS–842C, 5′-AGACGGGGACTCCCGAGC-3′; TS–30A, 5′-CTGCTTTTATTCTCTAATCTGCTCCCA-3′; TS–30T, 5′-CTGCTTTTATTCTCTAATCTGCTCCCT-3′; TS–30TS, 5′-CTTTTATTCTCTAATCTGCTCCCT-3′; TS–13AR, 5′-ACCCTCTGAAAGTGGACCA-3′.

Primers for telomere mapping

TAG-TELW, 5′-**TCATGCGTCCATGGTCCGGA**CCCTTACCCTTACCCTNACCCTA-3′; TAG-TELX, 5′-**TCATGCGTCCATGGTCCGGA**CCCTTACCCTTACCCTNACCCTC-3′; TAG-TELY, 5′-**TCATGCGTCCATGGTCCGGA**CCCTTACCCTTACCCTNACCCTG-3′.

Sequencing

Genomic DNA (100 ng) was amplified with primers TSK8G + TSK8C, TSK8E + TSK8B, TSK8J + TSK8K or TSK8E + TELG in 20 μl and the reactions were cycled 32 times at 96°C for 40 s, 65°C for 50 s and 70°C for 1 min. The products were purified by electrophoresis and electroelution onto a dialysis membrane, and the double-stranded products were sequenced with one of the PCR primers (Winship, 1989). A total of 803 bases were sequenced by this strategy. For the identification of the –13 polymorphism, PCR products (~500 bp) after amplification with TSK8E + TAG-TELX were sequenced with TSK8A. Allele-specific amplification from DNAs heterozygous for A/B flanking haplotypes, prior to the sequencing of B or C haplotypes, was achieved by the amplification of genomic DNA with TSK8B to TSK8C at 96°C for 20 s and 70°C for 140 s for 33 cycles. Allele-specific amplification was verified by digesting a portion of the amplified products with *Dde*I; the remaining products were sequenced with primers TSK8B, TSK8E, TSK8G or TSK8C.

RFLP analysis

Polymorphic positions –415, –427, –652 and –842 were assayed by the amplification of genomic DNA with primers TSK8G + TSK8C in 20 μl. In all, 5 μl of the PCR were digested with the restriction enzyme (Gibco-BRL) appropriate for the polymorphic position (see Table I) in a 16 μl reaction and in the presence of 1 mM spermidine trichloride. The products were resolved on a 2.2% Metaphor (FMC) agarose gel. RFLP analysis of the –176 position was achieved by the amplification of genomic DNA in the presence of TSK8E (1.0 μM), TSK8B (0.5 μM) and TS–13AR (0.5 μM). The products were digested to completion with *Dde*I and resolved on a 4.5% Metaphor agarose gel. The polymorphism at –1888 was assayed by the digestion of TSK8J + TSK8K PCR products with *Nla*IV.

ARMS assays for the –13 and –30 polymorphisms

The –13 polymorphism was identified during an analysis of the –176 and –30 polymorphisms because of the inefficient amplification by the TSK8B primer. The –13 polymorphism was assayed using the allele-specific primer TS–13AR with TSK8E. PCRs in the presence of TS–13AR were cycled 32 times at 96°C for 50 s, 66°C for 40 s and 72°C for 1 min. Products were resolved in 1.5% HGT agarose (FMC) gels. The –30 polymorphism was assayed by PCR (10 μl) of genomic DNA in the presence of 1.0 μM of the allele-specific primer TS–30A or TS–30TS in combination with 0.5 μM of both TSK8B and TS–13AR primers. Reactions in the presence of TS–30A were cycled 32 times at 96°C for 15 s, 68°C for 40 s and 70°C for 30 s; in the presence of TS–30T reactions were cycled 33 times at 96°C for 15 s, 62.5°C for 40 s and 70°C for 30 s. The products were resolved on a 4.5% Metaphor (FMC) agarose gel.

Haplotype analysis

From the –842 polymorphic position, amplification reactions (15 μl) with the allele-specific primers TS–842C and TSK8G were cycled 32 times at 96°C for 40 s, 67.5°C for 50 s and 70°C for 60 s. 4 μl of each PCR were digested with the appropriate enzyme for the polymorphic position (–415, –427 and –652) and resolved as above.

From the –13 polymorphic position, amplification reactions (10 μl) in the presence of the allele-specific primers TSK8B or TS–13AR with TSK8C or TSK8J were cycled as described in the –13 ARMS assay but the extension times were increased to 120 s at 72°C or 140 s at 72°C for the TSK8C and TSK8J primers, respectively. To type the –176 and –842 polymorphisms, the TSK8B/TS–13AR + TSK8C PCR products were digested with *Dde*I and were resolved in 3.5% Metaphor agarose gels; to establish haplotypes with the –1888 polymorphism, the TSK8B/TS–13AR + TSK8J PCR products were digested with *Nla*IV and resolved on a 2% HGT agarose gel.

TVR mapping

The allele-specific primer TS–30A or TS–30T was 5′ end-labelled in a 10 μl reaction containing 0.5 μM primer, 50 mM Tris–HCl (pH 7.6), 10 mM $MgCl_2$, 2 U T4 polynucleotide kinase and 370 kBq [γ-^{32}P]ATP (Amersham International) at 37°C overnight. Genomic DNAs (1 μg) selected for telomere mapping were digested with *Alu*I and *Dde*I in a 10 μl reaction for 2 h at 37°C. These enzymes do not cut between the TS–30A/T priming site and the Xp/Yp telomere repeat array, and were used to eliminate the low level of 'background' products from mispriming by TS–30A or TS–30T. The digested DNA (100 ng) was used in each of three PCRs (10 μl) containing 2 μl of the labelled primer and 0.4 μM of TAG-TELW, TAG-TELX or TAG-TELY telomere repeat primers. Reactions in the presence of TAG-TELW and TAG-TELY were cycled 19 times at 96°C for 20 s, 68°C for 40 s and 70°C for 2 min; reactions in the presence of TAG-TELX were cycled 19 times at 96°C for 20 s, 67.5°C for 40 s and 70°C for 2 min. The products (6.5 μl) were resolved by electrophoresis in a 6% denaturing polyacrylamide 0.5–2.5× TBE buffer gradient gel containing 7.67 M urea; 10× TBE contains 0.02 M EDTA in 0.9 M Tris-borate at pH 8.3. 5% polyacrylamide extension gels containing 1× TBE and 7.67 M urea were used to resolve longer products. The products were detected by autoradiography.

Acknowledgements

This work is the subject of a patent application. We thank Yuri Dubrova, John Armour, Celia May, Joanna Coleman, Andy Collick, David Neil and Meran Owen for helpful discussions and comments on the manuscript. N.J.R. is a UK HGMP Senior Fellow. This work has been supported by grants from the MRC and in part by an International Research Scholars award to A.J.J. from the Howard Hughes Medical Institute.

References

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

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.

Biessmann,H. and Mason,J.M. (1993) Genetics and molecular biology of telomeres. *Adv. Genet.*, **30**, 185–249.

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

Brown,W.R.A., MacKinon,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.

Buard,J. and Vergnaud,G. (1994) Complex recombination events at the hypermutable minisatellite CEB1 (D2S90). *EMBO J.*, **13**, 3203–3210.

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

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.

Decorte,R., Wu,R., Marynen,P. and Cassiman,J.-J. (1994) Identification of internal variation in the pseudoautosomal VNTR DXYS17, with nonrandom distribution of the alleles on the X and the Y chromosomes. *Am. J. Hum. Genet.*, **54**, 506–515.

Ellis,N., Kidd,J., Goodfellow,P.J., Kidd,K. and Goodfellow,P.N. (1990) Strong linkage disequilibrium between the *XY274* polmorphism and the pseudoautosomal boundary. *Am. J. Hum. Genet.*, **46**, 950–955.

Greider,C.W. (1994) Mammalian telomere dynamics: healing, fragmentation shortening and stabilization. *Curr. Opin. Genet. Dev.*, **4**, 203–211.

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.

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.

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

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.

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.A.L. (1994) Complex gene conversion events in germline mutation at human minisatellites. *Nature Genet.*, **6**, 136–145.

Kim,N.W. *et al.* (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.

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.

Monckton,D.G., Tamaki,K., MacLeod,A., Neil,D.L. and Jeffreys,A.J. (1993) Allele-specific MVR–PCR analysis at minisatellite D1S8. *Hum. Mol. Genet.*, **2**, 513–519.

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

Neil,D.L. and Jeffreys,A.J. (1993) Digital typing at a second 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). *Nucleic Acids Res.*, **17**, 2503–2516.

Nickerson,D.A., Whitehurst,C., Boysen,C., Charmley,P., Kaiser,R. and Hood,L. (1992) Identification of clusters of biallelic polymorphic sequence-tagged sites (pSTSs) that generate highly informative and automatable markers for genetic linkage mapping. *Genomics*, **12**, 377–387.

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.

Rappold,G.A. (1993) The pseudoautosomal regions of the human sex chromosomes. *Hum. Genet.*, **92**, 315–324.

Royle,N.J., Clarkson,R.E., Wong,Z. and Jeffreys,A.J. (1988) Clustering of hypervariable minisatellites in 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. Roy. Soc. Lond., Ser. B*, **247**, 57–61.

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.

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.

Simmler,M.-C., Rouyer,F., Vergnaud,G., Nystrom-Lahti,M., Ngo,K.Y., de la Chapelle,A. and Weissenbach,J. (1985) Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. *Nature*, **317**, 692–697.

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.

Weir,B.S. (1990) *Genetic Data Analysis*. Sinauer Associates Inc., Sunderland, MA 01375, USA.

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

Received on June 12, 1995; revised on July 20, 1995