The plasticity of human telomeres demonstrated by a hypervariable telomere repeat array that is located on some copies of 16p and 16q

Joanna Coleman, Duncan M. Baird and Nicola J. Royle*

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

Received September 23, 1998; Revised and Accepted June 8, 1999

Human telomeres are composed of tandem arrays of TTAGGG repeats with many variant repeats at the proximal ends. Comparison of the interspersion of variant and TTAGGG repeats between alleles can be used to study telomere instability, but the difficulty in identifying chromosome-specific sequences close to the start of autosomal telomeres has hampered such investigations. A chromosome end, including a telomere and adjacent sequence, that is polymorphic for its presence or absence in unrelated individuals has been identified. The telomere-adjacent DNA shows strong homology (92–99%) to sequences, including two expressed sequence tags, that are usually located in subterminal regions of human chromosomes but not adjacent to telomeres. Since this chromosome end arose, it has relocated at least once. In Caucasians, it forms the telomere of ~6% of 16q and 2% of 16p chromosome arms. The mechanism of relocation is unknown but must have involved the telomere-adjacent DNA rather than the telomere itself, as copies on 16p and 16q share the same telomere-adjacent sequence. The interspersion patterns of TTAGGG with TGAGGG, TTGGGG and non-amplifying repeat sequences revealed extensive allelic variation, such that 47 different alleles were observed among the 50 alleles mapped. Closely related alleles differ by small changes in copy number at blocks of adjacent like repeats, as seen at the Xp/Yp pseudoautosomal telomere. Such differences are compatible with a model in which the majority of mutations arise by intra-allelic mechanisms, in individuals hemizygous for a single copy of the chromosome end.

INTRODUCTION

The evolution of terminal sequences in higher primates has been rapid, for example a subterminal satellite has expanded considerably and dispersed to many chromosome ends in the chimpanzee and gorilla lineages but is absent from the human genome (1). The mechanisms underlying the rapid turnover of terminal sequences in primates are likely to include terminal deletions, unequal exchanges between subterminal sequences of homologous chromosomes and exchanges between subterminal sequences of non-homologous chromosomes (2). Such rearrangements may be a normal feature of the dynamic subterminal regions and telomeres of human chromosomes; however, errors in the normal turnover processes could give rise to an unbalanced genome. Many patients with unbalanced genomes arising from rearrangements of terminal sequences have been described (see, for example, refs 3–5). In order to understand what events give rise to such rearrangements and why they occur, it is important to know what processes turn over subterminal sequences and telomeres at normal chromosome ends.

Sequence analysis of subtelomeric regions of several human chromosomes has revealed some similarities in organization (6). The subtelomeric regions can be subdivided into proximal and distal regions, sometimes separated by a short degenerate tandem array of TTAGGG repeats (7,8). The proximal subtelomeric regions contain long stretches of sequences that show homology to a few other chromosome ends, whereas the distal subtelomeric regions contain a more complex patchwork of sequences with homology to many chromosome ends (6). In addition, sequences of the distal telomeric regions of 16p, 4p, 22q and 17p and others show homology to many expressed sequence tags (ESTs) (6,9). To some extent, the sequence organization resembles that seen in Saccharomyces cerevisiae, and it has been suggested that proximal and distal subtelomeric sequences may interact differently with other non-homologous chromosomes (10).

In human chromosomes, subtelomeric regions are terminated abruptly by the presence of a telomere which is composed of a tandem array of a short repeat unit. The sequence of human telomeres is determined, primarily, by the enzyme telomerase, which adds TTAGGG repeats onto the terminus in the germline, so maintaining the length of the telomere for the next generation (11). The distance between the start of the telomere repeat array and its terminus varies considerably, but it can be as much as 20 kb. Repeats at the proximal end of the array are rarely influenced by the activity of telomerase and may accumulate mutations and diverge from the original sequence. However, telomeres are also important for the protection of chromosome integrity, and the interaction of the telomere repeat array with proteins, such as TRF1 and TRF2, may place constraints on the turnover of repeats at the proximal end of the repeat array (12,13).

*To whom correspondence should be addressed. Tel: +44 116 2522270; Fax: +44 116 2523378; Email: njr@le.ac.uk
Detailed analysis of the Xp/Yp pseudoautosomal telomere revealed extensive allelic variation at the proximal end of the repeat array itself, such that among 65 Caucasian alleles analysed, 61 different alleles were observed (14). The high level of allelic diversity indicated that there must be a high underlying mutation rate. Comparison of the interspersion patterns of TTAGGG and variant repeat types (TGAGGG and TCAGGG) showed that alleles could be subdivided into groups of related alleles that shared similar interspersion patterns. Alleles within a group tended to differ by small changes in the copy number of like repeats in a block. Also, related alleles were associated with only one haplotype in the DNA immediately adjacent to the Xp/Yp telomere. These features suggest that mutations underlying the allelic diversity are dominated by intra-allelic events such as replication slippage or unequal sister chromatid exchange. As a consequence, the Xp/Yp telomere repeat arrays appear to have evolved along haploid lineages with few exchanges between alleles. This surprising observation might be an unusual feature confined to the Xp/Yp telomere and adjacent sequences. The Xp/Yp telomere is distinct from other telomeres in the genome, in that it is present at the end of the pseudoautosomal region 1 (PAR1) and the subtelomeric sequences at Xp/Yp are mostly unique. Autosomal telomeres may be influenced by different factors compared with the Xp/Yp telomere and adjacent sequences and, therefore, it is essential that individual autosomal telomeres are investigated if the significance of the Xp/Yp data is to be understood. In order to address these questions, we have analysed sequence variation at the proximal end of a polymorphic autosomal telomere found on human chromosome 16.

RESULTS

The clone Nitul14e was isolated from a library of telomere junction clones generated by telomere-anchored PCR (15) from a human DNA (Nitu). This and several identical but independently isolated clones contained 210 bp of DNA extending from an MboI restriction site to an array of TTAGGG and variant repeat types (Fig. 1). The DNA adjacent to the repeat array contains a short palindromic sequence only 5 bp from the first TTAGGG repeat. The entire DNA sequence adjacent to the repeat array shows strong homology to several sequences in the DDBJ/EMBL/GenBank databases. Two of these sequences are uncharacterized ESTs (98% homology to AA016323, Fig. 1; 97% homology to H09685). Others are found in the subtelomeric regions of human chromosomes as follows: 4p (94% homology to Z95704), 4q (95% homology to AF035190), 21q (96% homology to AC002055, cosmid n1g3 within 50 kb of the telomere). In addition, there are two matches to sequenced clones of unknown locations (99% to AC004908 and 93% to AC004842), but provisional analysis indicates that these clones may be from subterminal regions. Thus the Nitul14e clone contains a copy of a sequence family that is often located near the ends of human chromosomes and some of which may be transcribed. In order to investigate the copy number and distribution of the sequence further, PCR analysis of a monochromosome hybrid panel was carried out using primers Nitul14eA and Nitul14eB. A product of uniform size but variable intensity was amplified from the majority of...

---

*Figure 1.* Comparison of the sequence of the Nitul14e clone with the EST, AA016323. The sequence of the Nitul14e clone is shown from the Sau3AI site (GATC) at the 5' end and includes the first three telomere repeats at the 3' end (italics). It is aligned with the EST sequence (AA016323) with which it shows 98% homology. Vertical lines show identical bases between the two sequences; asterisks represent mismatches. The positions of the primers are shown as solid arrows. The short imperfect 5 bp palindrome is shown by inverted dashed arrows. The sequence is numbered from the start of the telomere repeat array; -1 is the first telomere-adjacent base (accession no. AJ011097).
the human chromosomes, with the exception of the sex chromosomes and autosomes 11, 12 and 17 (data not shown), indicating that copies of this sequence family are found on most chromosomes.

Identification of a polymorphic telomere

The unusual feature of the Nitu14e telomere junction clone was the truncation of the EST-like sequence by an array of TTAGGG and variant repeats. None of the homologous sequences in the databases are terminated by TTAGGG and variant repeat arrays. In order to verify the sequence organization of the cloned DNA, PCR analysis from the flanking sequences with either primer Nitu14eA or Nitu14eC and a primer, TelC, that anneals to the G-rich strand of tandem TTAGGG repeats (15) was used to amplify genomic DNAs. The PCR products were resolved by agarose gel electrophoresis and the resulting Southern blot was hybridized to a probe (hereafter called the Nitu probe) generated from the Nitu14e clone using primers Nitu14eA and Nitu14eB. Among the nine Caucasian DNAs shown in Figure 2, two (CEPH DNA 1329301, 1329402 and Nitu) gave smears of strongly hybridizing fragments that extend from 0.2 kb upwards. The intensity of the hybridization signal (which faded out at ~2.5 kb), combined with the banding pattern superimposed on the smear, suggested that the products had arisen from the proximal end of a telomere that contains TTAGGG and variant repeat types. Discrete amplicons of ~1.5 kb were observed in five of the DNAs shown in Figure 2 (CEPH DNAs 1329301, 1329402 and weakly in 1329101, 1329202 and 1329401). These fragments are also polymorphic for their presence or absence but are independent of the smears and are therefore unlikely to be linked to them. Instead, they probably arise from another copy of the sequence family that is adjacent to a few internal TTAGGG repeats. The nature of this second polymorphic locus was not investigated.

To verify the terminal location of the polymorphic smear of hybridizing fragments, Bal31 sensitivity was examined. The Nitu DNA sample was treated with Bal31 for increasing lengths of time. The DNA was then subjected to a multiplex PCR for the simultaneous amplification of the polymorphic telomere (using Nitu14eC and TelC as seen in Fig. 2) and amplification of a 600 bp product from the D16S309 (MS205) locus [using primers 205H and 205J (16)]. The D16S309 locus is an estimated 1.3 Mb from the 16p telomere (16). In two separate experiments, Southern blot detection of the amplified products showed that the signal from the polymorphic smear became weaker and the longer fragments were lost with increasing Bal31 digestion. In contrast, the signal from the D16S309 locus did not decline with increasing exposure to Bal31 (data not shown). Given the sensitivity of the smear to
markers, and some families consistently gave negative lod scores, suggesting that there may be heterogeneity for linkage to the D16S7 locus among the CEPH families.

Evidence of heterogeneity for linkage to D16S7 was found among the families for which segregation data at D16S7 were available. Using the Homog program (17), the maximum likelihood for linkage heterogeneity (H2) exceeds that obtained for linkage without heterogeneity (H1); see Table 1. Further analysis using the Homog and linkage programs showed that, in CEPH families 17, 21, 1333, 1341, 1377, 13293 and probably family 2, the polymorphic telomere is at 16qter, but in families 1345 and 1375 it is located elsewhere. The mother in family 1340 carries one, paternally inherited copy of the polymorphic telomere at 16q and the other maternally inherited copy is elsewhere. With the knowledge that the polymorphic telomere segregates from two loci, locus 1 at 16q and locus 2 elsewhere, the two-point linkage analysis was repeated. Strong evidence was found for linkage between the polymorphic telomere locus 1 and the distal 16q loci: D16S7 (θ = 0.0, lod score = 10.8); D16S43 (θ = 0.034, lod score = 6.8) and D16S44 (θ = 0.0, lod score = 4.5). Linkage was also demonstrated between the polymorphic telomere locus 2 in the three families (CEPH families 1340, 1345 and 1375) and the distal 16p locus, D16S85 (θ = 0.0, lod score = 6.02) but not to terminally located markers tested on other autosomes.

In order to determine whether the polymorphic locus 2 (linked to 16p) represents the telomere from one of the characterized alleles of 16pter (19), four somatic cell hybrids each containing a different 16pter allele were screened (samples kindly provided by Dr D. Higgs). PCR analyses showed that each hybrid contained DNA sequences from adjacent to the minisatellite locus D16S309 (MS205) that, as indicated above, is located ~1.3 Mb from the 16p telomere (16). PCR amplification with the Nitu14eA primer and telomere repeat primer (Te1C), followed by Southern blot analysis using the Nitu probe, showed that none of the hybrids contained a copy of the polymorphic telomere. However, amplification with the primers Nitu14eA and Nitu14eC generated amplicons of the expected size from all four hybrids but not from the mouse control DNA (data not shown). In summary, the polymorphic telomere does not represent one of the previously characterized alleles at 16pter (A, B, C or D).

Table 1. Results of linkage heterogeneity testing between the polymorphic telomere and D16S7

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Max. lnL</th>
<th>α</th>
<th>θ</th>
</tr>
</thead>
<tbody>
<tr>
<td>H2, linkage heterogeneity</td>
<td>16.68</td>
<td>0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>H1, linkage homogeneity</td>
<td>8.17</td>
<td>(1)</td>
<td>0.2</td>
</tr>
<tr>
<td>H0, no linkage</td>
<td>(0)</td>
<td>(0)</td>
<td>(0.5)</td>
</tr>
</tbody>
</table>

Family no. | Conditional probability of linkage to D16S7 | Lower support limit | Upper support limit |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>13293</td>
<td>0.9869</td>
<td>0.9283</td>
<td>0.9984</td>
</tr>
<tr>
<td>1340</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0017</td>
</tr>
<tr>
<td>1341</td>
<td>0.9934</td>
<td>0.9627</td>
<td>0.9992</td>
</tr>
<tr>
<td>1375</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0062</td>
</tr>
<tr>
<td>1377</td>
<td>0.9983</td>
<td>0.9901</td>
<td>0.9998</td>
</tr>
<tr>
<td>21</td>
<td>0.9737</td>
<td>0.8664</td>
<td>0.9967</td>
</tr>
<tr>
<td>17</td>
<td>0.9967</td>
<td>0.9805</td>
<td>0.9996</td>
</tr>
</tbody>
</table>

α is the proportion of families showing linkage, θ is the recombination fraction.
Table 2. Occurrence of the polymorphic telomere in different populations

<table>
<thead>
<tr>
<th>Population</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEPH parents</td>
<td>13/160 (8.1%)</td>
</tr>
<tr>
<td>16q = 6.3% 16p = 1.9%</td>
<td></td>
</tr>
<tr>
<td>CEPH grandparents</td>
<td>13/210 (6.2%)</td>
</tr>
<tr>
<td>Other Caucasians</td>
<td>11/162 (6.8%)</td>
</tr>
<tr>
<td>Japanese</td>
<td>7/128 (5.5%)</td>
</tr>
<tr>
<td>Afro-Caribbean</td>
<td>6/134 (4.5%)</td>
</tr>
<tr>
<td>Asian</td>
<td>6/80 (7.5%)</td>
</tr>
<tr>
<td>Zimbabwean</td>
<td>8/94 (8.5%)</td>
</tr>
<tr>
<td>Malaysian</td>
<td>22/428 (5.1%)</td>
</tr>
</tbody>
</table>

The frequencies for the CEPH parents are allele frequencies, determined from segregation analysis. For all other populations, the frequency is only an estimate of the allele frequency (see text). The Asian population includes only individuals from the Indian subcontinent (mainly Bangladesh and Pakistan).

Presence of the polymorphic telomere in different populations

In order to determine the occurrence of the polymorphic telomere in different populations, a survey of unrelated genomic DNAs was undertaken. Table 2 shows that the polymorphic telomere is present at a low frequency in each of the populations examined. The frequency in the CEPH parents is an allele frequency determined from segregation analysis. For the other populations, segregation analysis was not possible but, assuming that the presence of a banded smear represents a single allele in the majority of individuals, frequencies have been estimated. As a result, the frequencies shown in Table 2 are underestimates of the allele frequencies in these other populations. Without segregation analysis, it is not possible to determine the location of the polymorphic telomeres in these populations or indeed the number of locations.

Variation in the polymorphic telomere repeat arrays

As noted above, differences in the banded smears highlighted further variation between alleles of the polymorphic telomere. To investigate the extent of this variation further, unrelated DNA samples positive for the telomere were compared (Fig. 4). The diffuse bands represent short blocks of TTAGGG repeat types interspersed with non-hybridizing segments that contain either variant repeat types or short sequences of non-repetitive DNA. The majority of the patterns analysed by this method were different, although some patterns were clearly more similar than others (e.g. Japanese samples J446, J462, J428 and J443, Fig. 4). The interspersion patterns of TTAGGG repeats with other unknown sequences extends for at least 1.5 kb in many of the polymorphic telomeres, but in a few (e.g. AS19, Fig. 4) the initial banding pattern was replaced by a continuous smear of longer TTAGGG hybridizing fragments.

Analysis of the interspersion patterns in the polymorphic telomere repeat array by TVR–PCR

The full extent of the allelic variation cannot be detected using low resolution agarose gel electrophoresis and a single primer that detects only TTAGGG repeat types. Therefore, the interspersion of TTAGGG with TGAGGG and TTGGGG repeat types was examined using telomere variant repeat mapping (TVR–PCR) (14). As the majority of DNA samples positive for a polymorphic telomere smear are likely to be hemizygous and carry only a single copy of the telomere, TVR–PCR analysis was conducted using the telomere-adjacent primer Nitu14eD that will amplify from any copy of the polymorphic telomere present in the DNA sample. The Nitu14eD primer, in combination with one of the telomere primers TAGTelW, TAGTelX or TAGTelJ (that amplify from TTAGGG, TGAGGG and TTGGGG repeat types, respectively), was used to generate high resolution maps of the individual alleles. Figure 5 shows examples of the telomere maps generated from five CEPH DNAs; three of them (1701, 2101, 133301) are hemizygous for a polymorphic telomere at 16q, one (137502) is hemizygous for a polymorphic telomere at 16p and the last (134002) is a compound hemizygote with one copy on 16q and one on 16p. The telomere maps of the four hemizygous individuals are simple, with bands appearing in only one or none of the three tracks (T, G or J) at each position of the 6 bp ladder. These patterns can be converted relatively easily into a code that defines the interspersion of T-, G- or J-type repeats with non-amplifying or null sequences. The pattern seen in the com-
A total of 50 alleles was mapped using TVR–PCR and 47 different codes were observed, demonstrating the hypervariable nature of the proximal end of these telomere repeat arrays (Fig. 6A). Most alleles contained several non-amplifying sequences; these were probably not TCAGGG repeat types (data not shown) but are most likely other variant repeats that do not amplify with any of the primers used. Some telomere maps faded out after ~20 repeats (e.g. 137502 and R46), but inspection of the autoradiographs from agarose gels revealed that these telomeres have large gaps in the hybridization pattern (see R46 in Fig. 4). Three pairs of identical TVR codes were observed among CEPH DNAs, and comparison of the agarose gel patterns indicated that these pairs of alleles are identical for at least 1.5 kb into the array, though minor changes would not be detected by this method.

The majority of the telomeres started at the same position with respect to the flanking primer, with a run of T-type repeats. In the absence of any clear differences that could be used to subdivide the alleles into groups, the allele codes have been aligned from the first repeat unit and arbitrarily subdivided into groups according to the number of T-type repeats at the start of the array. Small groups of closely related alleles do, however, share very similar interspersion patterns, e.g. Jebr, 141181, 1329302, Chha or J443, J428, J446 and J461. As observed at the Xp/Yp telomere, the variation between alleles was mostly because of small changes in the copy number of like repeat types.

An attempt was made to identify sequence differences that might be used to distinguish polymorphic telomeres located at 16q from those probably located at 16p. TVR–PCR analysis of the 16q- and 16p-associated polymorphic telomeres was carried out using the Nitu14eA and TAGTelW primers. Bands corresponding to the first few repeats in the telomere repeat array were isolated individually and used as templates to sequence the telomere-adjacent DNA. Sequence comparison of 190 bp of the adjacent DNA did not show any differences between the 16q- and 16p-associated polymorphic telomeres. As a result, it was not possible to identify the location of most of the mapped alleles shown in Figure 6A. However, alleles that have been mapped by segregation analysis have been compared (Fig. 6B). Like the 16q alleles, the three 16p alleles start with a block of T-type repeats but each also has a run of J-type repeats after which the TVR maps stop because of a long stretch of non-amplifying sequence. Although suggestive of greater similarity between these alleles, there is insufficient data to draw any firm conclusions.

**DISCUSSION**

The polymorphic distribution of subterminal repeat sequences on human chromosomes suggested that length variation between the most distal chromosome-specific sequence and the telomere would be identified on some chromosomes (18). The only well-characterized example of such variation is at the terminus of 16p, where four different length alleles (A–D) have been identified (19), although length variation has been reported at other chromosome ends; for example, three different length alleles have been described at 16q (20).

We have identified a chromosome end that we have described as a polymorphic telomere because the telomere repeat array with the adjacent EST-like sequence can be present or absent from the genome in unrelated individuals. The length of the repeat arrays (>2.5 kb) makes it unlikely that it is an internal array of telomere-like repeats, as such arrays tend to be short (<1 kb) (6), and the Bal31 sensitivity of the repeat array confirms a terminal location. Linkage analysis in CEPH families indicates that this polymorphic telomere is located primarily on 16q.
The distribution of the EST-like sequences is of interest. PCR analysis of the monochromosome hybrid panel and four somatic cell hybrids, containing different copies of chromosome 16, shows that the EST-like sequence is present in at least one copy on many different chromosomes. In addition, the homologous sequences identified in sequence databases map to terminal regions of human chromosomes (e.g. AL035696 at 6p25), or have a structure similar to sequences from subterminal regions (AC004842 and AC004908). This suggests that the majority of copies of this family of EST-like sequences are

Figure 6. Alignment of telomere codes from the polymorphic telomere. (A) Telomere codes for 50 alleles are presented. They are aligned at the left-hand side from the first repeat in the array. The alleles have also been grouped according the number of T-type repeats shown at the start of the array (see text). (B) Telomere codes for the 11 alleles present in the CEPH panel have been grouped by chromosome location. The DNA sample code and ethnic origin are shown on the left, as is the chromosome location when known. The end of the code indicates either a long stretch of non-amplifying DNA or the limit of the 6 bp ladder that could be coded reliably. Pairs of identical codes are shown in italics. Ethnic origins are Zi, Zimbabwean; Ca, Caucasian; J, Japanese; Ma, Malaysian; AS, Asians from the Indian subcontinent. The allele codes are composed of TTAGGG (T), TGAGGG (G) and TTGGGG (J) repeat types. Non-amplifying sequences (N) that are most likely other variant repeat types are displayed as dashes (–). The sequence composition of the non-amplifying sequence (.) at the start of an array is unknown.
located in subterminal regions. At least one of these sequences (located in cosmid n1g3, AC002055) lies within 50 kb of the 22q telomere, but from analysis of telomere-anchored PCR libraries from unrelated individuals it is clear that this EST-like sequence is not usually found immediately adjacent to telomeres (15; N.J. Royle and H. Varley, unpublished data).

Furthermore, the EST-like sequence is not a feature of all subterminal regions because the shortest allele of 16pter (allele A) does not contain a copy.

It is interesting to speculate how, when and where the polymorphic telomere may have arisen. Telomerase-mediated healing of a chromosome break at a subterminal copy of the EST-like sequence could have given rise to a telomere at a novel location. Alternatively, it could have arisen as an interstitial deletion moving a pre-existing telomere to a more proximal location. The presence of the polymorphic telomere in several populations and the considerable variation between the alleles suggest that it has an ancient origin, probably pre-dating the radiation of modern man from Africa (21,22).

The chromosomal origin of the polymorphic telomere is unknown because of the abundance of this EST-like sequence in the modern genome. Subsequent to the formation of the telomere adjacent to the EST-like sequence, there has been at least one duplication event that has resulted in a copy of the telomere and adjacent DNA being located at 16q (6% of Caucasians) and at 16pter (2% of Caucasians). The copy of the polymorphic telomere located at 16q may represent one of the three 16q alleles identified previously (20). The event that resulted in the telomere and adjacent EST-like sequence moving (jumping) from one end of chromosome 16 to the other must have involved the telomere-adjacent DNA rather than the telomere because the telomere repeat arrays start at the same base in each location. The limited sequence data (190 bp) that we have adjacent to the telomere repeat array show no differences between the 16q and 16p copies, so it is not possible to identify the mechanism involved.

Without segregation data, the location of the polymorphic telomere in the other populations examined is not certain. There is evidence that terminal sequences, perhaps including the telomere, have been exchanged between 4q and 10q during recent human evolution. However, such exchanges probably occur infrequently, as no clear examples of de novo mutations involving the 10q and 4q homologous sequences have been identified in families (24–26). If subterminal exchanges that involve the relocation of a telomere are infrequent, it is reasonable to assume that some copies of the polymorphic telomere described here are on chromosome 16 in all the populations we examined.

Comparison of 50 telomeres from the polymorphic telomere has revealed considerable variation between the alleles. The three pairs of identical alleles seen in Caucasians may indicate a lower level of variation compared with the Xp/Yp telomere. A mutation rate of $10^{-5}$ (0.1%) has been estimated for the polymorphic telomere in Caucasians, from the equation $4N_{e}u = \theta$ ($N_{e}$ is the effective population size, $\mu$ is the mutation rate and $\theta$ is the allelic diversity; the allelic diversity was determined under the infinite allele model from the observation that 23 different alleles were seen among 26 Caucasian alleles analysed).

The interspersions patterns of the closely related alleles tend to differ by small changes in copy number at blocks of adjacent like repeats. It is also easier to align the alleles from the start of the array, as the similarities become less clear further into the array.

Such differences between related alleles suggest that intra-allelic mutation mechanisms underlie the allelic diversity. Indeed, the fact that the polymorphic telomere occurs singly (hemizygous) in most individuals also suggests that intra-allelic mutations are more likely to underly the allelic diversity. However, opportunities for exchanges between two alleles at the same chromosome end do occur (e.g. ~1.1% of Caucasians are heterozygous for different copies of the polymorphic telomere at 16q) and cannot be excluded as a mutational mechanism contributing to allelic diversity. Furthermore, opportunities for exchanges between copies of the polymorphic telomere at 16p and 16q also occur. On balance, however, the data from the Xp/Yp telomere (14) and now these polymorphic telomeres of chromosome 16 indicate that telomere repeat arrays are evolving independently, along haploid lineages.

**MATERIALS AND METHODS**

**Genomic DNAs**

Caucasian DNA samples consisted of lymphoblastoid DNAs from the CEPH panel of 40 family DNAs; other Caucasian blood DNA samples were collected in Leicester over a number of years. The DNA samples from different ethnic groups were from the CEPH panel of 40 family DNAs; other Caucasian DNA samples consisted of lymphoblastoid DNAs from unrelated individuals it is clear that this EST-like sequence is not usually found immediately adjacent to telomeres (15; N.J. Royle and H. Varley, unpublished data). Furthermore, the EST-like sequence is not a feature of all subterminal regions because the shortest allele of 16pter (allele A) does not contain a copy.

Cloning telomere junction fragments

The method has been described elsewhere (15), but briefly it entails size selection of MboI-digested genomic DNA fragments $>5$ kb; ligation of the size-fractionated DNA to Sau3AI linker amplimers; and amplification of the pool of telomere-enriched fragments using a linker amplimer (Sau-L-A) and a primer (TelC) that anneals to the G-rich strand of human telomere repeat arrays. The amplified products were cloned into the EcoRI–KpnI sites of the pBluescript SKII+ vector. Telomere-positive clones were identified by hybridization to a (TTAGGG)$_n$ probe and single-stranded products were sequenced.

**PCR primer sequences**

Primers Sau-L-A and TelC have been reported before (15). Other primer sequences were as follows: Nitu14eA, 5'-GATCGAGCCAGGGTCTGATC-3'; Nitu14eB, 5'-GGAGACGCTCCTAGCTCAGAG-3'; Nitu14eC, GCTGTGATGCGGCTGCCACC-3'; and Nitu14eD, 5'-CTCTGAGTCAGGAGCTG-3'. Primers 205H and 205J (16), TAG (27) and TAGTelW and TAGTelX have been reported before (14). TAGTelI was 5'-TCTAGGCTCCATGTGCCGAACCCCCAACCACCCACCC-3'. The non-complementary TAG sequence is not usually found immediately adjacent to EST-like libraries from unrelated individuals it is clear that this EST-like sequence is not usually found immediately adjacent to telomeres (15; N.J. Royle and H. Varley, unpublished data). Furthermore, the EST-like sequence is not a feature of all subterminal regions because the shortest allele of 16pter (allele A) does not contain a copy.

DNA amplification

PCRs were carried out using the buffer system described previously (28) typically in a 10 µl volume using 50 ng of human genomic DNA, 1 µM of each primer and 0.1 U/µl Taq polymerase (Advanced Biotechnologies). Amplifications were...
carried out using a PTC-200 thermal cycler (MJ Research) or a GeneAmp 9600 thermal cycler (Perkin Elmer). The PCR cycling conditions to amplify the polymorphic telomere repeat array with flanking primers Nitu14eA or Nitu14eC and TelC were 96°C for 20 s, 69°C for 40 s and 70°C for 5 min for 20 cycles. The low cycle number avoids the formation of collapsed products that are generated at high cycle number. The Nitu probe was generated from the clone Nitu14e by PCR using primers Nitu14eA and Nitu14eB.

Detection of the polymorphic telomeres
The amplified products were resolved by gel electrophoresis typically using 1.2% agarose gels (FMC-HGT agarose) in a Tris–borate buffer system (16). The DNA was blotted onto Hybond-Nfp (Amersham International) and hybridized in phosphate-SDS solution at 65°C to the radioactively labelled Nitu probe; post-hybridization washes were used prior to autoradiography.

Bal31 sensitivity
High molecular weight Nitu DNA (12 μg) was digested in the presence of 12 U of Bal31 in the buffer recommended by the manufacturer at 30°C. Aliquots containing 1 μg of the genomic DNA were removed at 0, 30, 60, 90, 120 and 195 min time points. The reactions were stopped by the addition of an equal volume of 40 mM EGTA and heat treatment (65°C for 10 min). The DNA aliquots were purified by phenol–chloroform extraction and ethanol precipitation. DNA concentration was checked using a fluorimeter (Hoefer TK100) and equalized between aliquots. Equivalent quantities of Bal31-treated DNA for each time point were amplified in a multiplex PCR reaction for the amplification of the polymorphic telomere and the D16S309 locus. PCR reactions (10 μl) included primers Nitu14eC, TelC, 205H and 205J (each at 1 μM) and were cycled 20 times at 96°C for 20 s, 67°C for 40 s and 70°C for 2 min. The PCR products were divided into two and resolved by agarose gel electrophoresis. Southern blots were made and used to detect the polymorphic telomere or the D16S309 locus (as above). The relative intensity of the MS205:polymorphic telomere signal was determined from PhosphorImage (Molecular Dynamics) measurements, using the Image Quant version 3.3 software.

TVR–PCR
This was carried out as described before (14) with a few modifications. The γ-32P-end-labelled telomere-adjacent primer Nitu14eD was used in combination with either TAGTelW, TAGTelX or TAGTelJ (0.4 μM). Each 10 μl TVR–PCR contained 50 ng of genomic DNA and the PCRs were cycled 20 times at 96°C for 20 s, 67.5°C for 45 s and 70°C for 2 min. The denaturing acrylamide gel electrophoresis and autoradiography were carried out as before. For coding of the TVR maps, some rules were used to ensure reproducibility. In the T tracks at the start of the TVR maps, only bands of equal intensity were recorded as T-type repeats. Stutter bands were observed before the start and at the end of a block of J-type repeats; therefore, only bands from the first to the last strongest were counted as J-type repeats. N-type repeats were scored at rungs on the 6 bp ladder when there was no signal in all three tracks or at occasional positions of ambiguity.

Sequencing of acrylamide gel-purified DNA fragments
Small blocks of dried acrylamide gel containing the DNA fragment to be sequenced were soaked in 15 μl of distilled water at room temperature for 30 min. Aliquots (1 μl) were re-amplified in a 30 μl volume using primers Nitu14eA and TAG. The products (2 μl) were checked for size and concentration by agarose gel electrophoresis (4% NuSieve; FMC). Then 20–50 ng of purified double-stranded DNA fragments were sequenced using Applied Biosystems dye terminator sequencing chemistries. The sequences were analysed using the Applied Biosystems Autoassembler program (Version 1.4).

ACKNOWLEDGEMENTS
We wish to thank Dr Jonathan Flint for the Nitu lymphoblastoid cell line and Professor Sir Alec J. Jeffreys for the generous supply of DNA samples from different ethnic origins, for the CEPH family panel of DNAs and for valuable comments on the manuscript. Our thanks to Celia May for primers and advice on the D16309 locus and also to Helen Varley and all members of Labs G18 and G19 for their useful comments. Finally thanks to the UK-MRC, which funded the research.

REFERENCES