Sequences from higher primates orthologous to the human Xp/Yp telomere junction region reveal gross rearrangements and high levels of divergence

D. M. Baird and N. J. Royle*

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

Received August 7, 1997; Revised and Accepted September 17, 1997 DDBJ/EMBL/GenBank accession nos: Z96949, Z96950, Z96951, Z96952, Z96953

A high level of sequence polymorphism combined with linkage disequilibrium has created a limited number of highly diverged haplotypes across the human Xp/Yp telomere junction region. To gain insight into the unusual genetic characteristics of this region, we have examined the orthologous sequences in the common chimpanzee (Pan troglodytes), the gorilla (Gorilla gorilla) and the orang-utan (Pongo pygmaeus). Divergence from the human Xp/Yp sequence is higher (average 2.6-fold) than that observed at other loci. The position of the human Xp/Yp telomere is unique, as additional sequences are present at this location in the other three species. These included an array of sub-terminal satellite in the chimpanzee and, in the gorilla a small interstitial array of telomere-like repeats followed by sequences with strong homology to the human 18p subterminal region. In the orang-utan, two alleles with different structures were identified. These differ by the presence or absence of a short interspersed nuclear element (SINE) sequence just proximal to long arrays of telomere-like repeat sequences that probably represent the proximal end of the orang-utan Xp/Yp telomere. In addition, a high level of sequence divergence between the two orang-utan structures was identified. This divergence is similar to that observed between the human Xp/Yp telomere-adjacent haplotypes. The high sequence divergence and evidence of gross rearrangements indicate that the Xp/Yp telomeric region has evolved faster than the rest of the genome.

INTRODUCTION

In contrast to sequences adjacent to many autosomal telomeres, sequences adjacent to the human Xp/Yp telomere are largely unique in the genome (1). The terminal 1 kb of DNA immediately adjacent to the Xp/Yp telomere contains a unique monomorphic minisatellite composed of four 64 bp repeat units. In addition, a truncated copy of a short interspersed nuclear element (SINE) is located 14 bp from the start of the Xp/Yp telomere repeat array (1,2; Fig. 1). There are an estimated 20 000 copies of this family of SINES within the human genome (3).

Previously, we reported the characterisation of human sequence polymorphism in the DNA immediately adjacent to the Xp/Yp telomere (4). Strong linkage disequilibrium across 18 base substitutional polymorphisms and one insertion/deletion polymorphism were identified in the 850 bp adjacent to the human Xp/Yp telomere, creating three highly diverged haplotypes, termed A, B and C. Furthermore, linkage disequilibrium extends into the proximal end of the Xp/Yp telomere (4). The intermediate haplotypes generated by sequential mutation and/or recombination of this DNA segment appear to be absent from modern human populations, and so the processes underlying the generation and maintenance of such divergent human haplotypes are unclear. The maximum sequence divergence between existing haplotypes is 1.6% in Caucasians and 2.2% in Africans, similar to the divergence between chimpanzees and humans (4,5). This raises two possibilities, firstly that the haplotypes adjacent to the human Xp/Yp telomere are ancient, possibly older than the separation of the Pan and Homo lineages, or secondly that they have undergone rapid evolution.

Some data have indicated that the ends of human chromosomes are dynamic regions of the genome. For example, human subtelomeric repeat sequences are found at many but not all chromosome ends, and the distribution of these repeats is polymorphic between unrelated individuals (1,6–8). Detailed analysis of terminal sequences from human 16p has revealed extraordinary length variation to the start of the telomere, and four different length 16pter alleles have been identified. Intriguingly, some of the 16pter alleles showed more homology to telomere-adjacent sequences on other chromosomes than to other 16pter alleles, suggesting that exchanges between terminal sequences of non-homologous chromosomes ends have occurred (9,10). In addition, there is some evidence that subterminal sequences have been exchanged between 4q and 10q (11). The dynamic nature of subtelomeric regions is not confined to humans; for example, the sequence organisation in subtelomeric regions of Saccharomyces cerevisiae varies considerably between closely related species (12,13).

Previously, we reported the characterisation of telomere junction fragments from the chimpanzee genome (14). None of the sequences obtained showed any similarity to human telomere junction sequences. Long arrays of a tandemly repeated 32 bp...
AT-rich sequence were identified at about half the chimpanzee and nearly all the gorilla chromosome ends. However, this subterminal satellite is absent from the human and orang-utan genomes. These data suggest that telomeres have been ‘repositioned’ during the evolution of the closely related primates and that sequence organisation at chromosome ends is likely to be species specific. In order to investigate the evolutionary dynamics of a single primate telomere, we have characterised Xp/Yp subterminal sequences in the chimpanzee, gorilla and orang-utan genomes.

RESULTS

Clustering of sequence polymorphism in the terminal 1 kb adjacent to the human Xp/Yp telomere

The sequence analysis of the human Xp/Yp terminal haplotypes (A, B and C) was extended to 2 kb from the start of the telomere repeat array. Six alleles each of the Caucasian haplotypes A and B were sequenced from individuals homozygous for these haplotypes. Three haplotype C alleles were sequenced after allele-specific amplification from individuals heterozygous for haplotypes A and C. One additional polymorphism was identified at position –1018 from the start of the telomere. The polymorphism consisted of a C→A transversion, with A present in all haplotype A alleles and C in all haplotype B and C alleles sequenced. Therefore, based on the sequence analysis of 15 alleles, the –1018 polymorphism is in complete linkage disequilibrium with the Xp/Yp terminal haplotypes. Only one other polymorphism (at –1888) has been identified in the penultimate kilobase of sequence compared with the 13 Caucasian polymorphisms found in the kilobase of DNA immediately adjacent to the Xp/Yp telomere (4). This represents a significant clustering within the terminal 1 kb ($P < 0.05 \chi^2$, Fig. 1). The biological significance, if any, of the clustered polymorphisms and the strong linkage disequilibrium adjacent to the telomere is unknown.

Distance from DXYS14 to the telomere

The DXYS14 locus (detected by the probe 29C1) is a hypervariable minisatellite that has been used to define the distal end of the genetic and physical maps of the human Xp/Yp pseudoautosomal region (15,16). A distance of 3.6 kb between DXYS14 and the telomere was estimated from sequence analysis and PCR amplifications using a primer distal to DXYS14 (29c1B) with a primer adjacent to the telomere (either TSK8G or TSK8K; Fig. 1). The short arm pseudoautosomal region has been well preserved during primate evolution, and this includes some repetitive sequences, for example, DXYS2 (17). As in humans, the chimpanzee DXYS14-like sequences are located at the distal end of Xp and Yp. The chimpanzee locus is polymorphic and displays a pseudoautosomal mode of inheritance (18). To determine whether the DXYS14-like sequences also exist in the gorilla and the orang-utan, the 29C1 probe was hybridised to PstI-digested genomic DNAs. The presence of these sequences was confirmed as strongly hybridising fragments were identified in each species (Fig. 2). The orang-utan DXYS14-like locus is probably unique in the genome, and it may be monomorphic or have reduced variability as only a single hybridising fragment was detected in the two orang-utans analysed. Two or more hybridising fragments were detected in humans, chimpanzee and gorillas, suggesting that the DXYS14-like locus is also variable in chimpanzees and gorillas. The distance from the distal end of the DXYS14 repeat array to a position orthologous to the start of the human telomere was determined from sequence data and a combination of PCR amplifications. After removal of length variation due to the presence of minisatellite sequences, the distance is 5.0 kb in chimpanzees and gorillas. In one orang-utan individual, two lengths were measured, 5.0 and 5.3 kb, and these have been designated orL and orU respectively.

The close physical proximity of the DXYS14-like sequences in the chimpanzee, gorilla and orang-utan to sequences homologous to the human Xp/Yp telomere-adjacent sequences indicates that we have compared equivalent sequences in the inter-species comparisons made below. As it has been shown that the DXYS14-like sequence is located at the end of Xp/Yp in the chimpanzee, we can...
be certain that, in this species at least, the sequences analysed are in the same location as in the human genome.

The position of the Xp/Yp telomere in primates is lineage specific

Southern blot analysis of the primate DNAs indicated that the human telomere-adjacent minisatellite was not located on the terminal Sau3AI restriction fragment in chimpanzees, gorillas or orang-utans (data not shown). To investigate the terminal sequence organisation further, a primer (TSK8S) known to anneal to chimpanzee, gorilla and orang-utan DNA was used in PCR reactions in conjunction with a primer for TTAGGG sequences [TAG-TeIW, (4)]. Amplification products were detected by hybridisation to a probe for the human telomere-adjacent minisatellite (see Fig. 1). A smear of amplified products was detected in human DNA, indicative of amplification into the Xp/Yp telomere. A similar smear of amplified products was detected in the orang-utan, but the shortest orang-utan product was shorter than that observed in the human DNA (Fig. 3). No amplification products were detected in the chimpanzee, and in the gorilla only a short array of interstitial telomere-like repeats was detected. These data strongly suggest that the Xp/Yp telomere in chimpanzees, gorillas and orang-utans does not start in the same position as in humans.

The chimpanzee- and gorilla-specific subterminal satellite is located at many chromosome ends in both species (14). PCR amplifications were carried out to determine whether the additional Xp/Yp sequences in the chimpanzee and gorilla genomes are composed of arrays of subterminal satellite. A Southern blot of products amplified by a subterminal satellite primer (Cht7B) and TSK8S was hybridised to the human telomere-adjacent minisatellite. A smear of hybridising fragments was detected in the chimpanzee, but no hybridising fragments were detected in amplifications of human, gorilla or orang-utan DNA (Fig. 3). Clearly the chimpanzee contains an array of subterminal satellite in place of the Xp/Yp telomere found in humans.

Analysis of the sequence orthologous to the human Xp/Yp telomere-adjacent DNA

Amplifications of the primate DNAs were carried out using primers designed from the 2 kb of sequence adjacent to the human Xp/Yp telomere and sequences distal to DXYS14 (primer 29C1B). The primer positions are depicted in Figure 1, and primer sequences not reported before (1) are detailed in Materials and Methods. As a consequence of sequence variation between species, not all the PCR primer combinations resulted in amplification. Products from successful amplifications were sequenced and, where necessary, additional primers designed to fill the gaps. In addition, vectorette PCR (19) was employed in the gorilla to obtain distal sequences that could not be amplified with the available human primers (see Materials and Methods). Distal sequences from the chimpanzee genome were amplified for sequence analysis by excision of 5′ radioactively labelled PCR fragments generated with primers PrimA and Ch7B from denaturing polyacrylamide gels followed by reamplification with the same primers. Using these strategies, contiguous sequences were established from the chimpanzee, gorilla and orang-utan.

Chimpanzee

A total of 2119 bp of contiguous sequence was obtained from the chimpanzee genome (EMBL accession no. Z96949). It contains a 45 bp duplication of a sequence that is only present as a monomer in the human DNA (Fig. 4). The four repeat unit minisatellite is present, and there was no evidence of length variation among the three chimpanzees analysed. As indicated above, an array of subterminal satellite repeats started at a position equivalent to ~89 from the start of the human telomere (Fig. 4). Analysis of the size of the subterminal satellite repeat array from pulsed-field gel (PFG) Southern blots revealed DNA fragments of different lengths in NcoI digests from unrelated
chimpanzees (data not shown). This suggests that there is sequence variation within the subterminal satellite array and, as a result, it has not been possible to determine the length of the subterminal satellite, but it probably exceeds 27 kb.

**Gorilla**

The ‘telomere-adjacent’ minisatellite is present in the gorilla, with two alleles of three and four repeat units present among six individuals examined. Four of the gorillas were homozygous for the three repeat unit allele and two were heterozygous. A total of 2223 bp of contiguous sequence was obtained from the gorilla allele containing three minisatellite repeats (EMBL accession no. Z96950). Sequence heterogeneity within the minisatellite reveals that the third most distal repeat unit is absent (Fig. 4). The sequence analysis defined the block of interstitial telomere-like repeats. It is variable in length, as arrays of nine and 10 telomere-like repeats were present in the individual analysed. The start of the block of telomere-like repeats corresponds to the position –125 from the start of the human telomere (Fig. 3). Surprisingly the gorilla telomere-like repeat array and sequences beyond it have strong homology to subtelomeric DNA from human chromosomes 4p, 18p and 4q, (94, 93 and 92% respectively) (20–22). The 18p and 4q sequences contain interstitial telomere repeats in the same position as those in the gorilla Xp/Yp sequence.

DNA sequence from the gorilla allele that contained a four repeat unit minisatellite was also established over 2287 bp (EMBL accession no. Z96951) via allele-specific amplification with primers Gor4A and Gor4B (Fig. 4). Comparison of the sequence obtained from the two gorilla alleles (EMBL accession nos Z96950 and Z96951) revealed 11 base substitutional differences between them (sequence divergence of 0.4%). The frequency of base substitutional sequence variation between the two gorilla alleles (1 per 203 bp, see Fig. 4) is much lower than observed in the kilobase of sequence adjacent to the human Xp/Yp telomere.
The sequence differences between the species were distributed across the region, and between the two orang-utan structures. Comparisons were made and divergence estimates obtained using Bestfit (Genetics Computer Group version 6.2). Previously published sequence divergence estimates from the \( \varphi \eta \)-globin gene region were included for comparison (5). hu, human haplotype A, ch, chimpanzee, gor, gorilla and orU/orL, orang-utan upper and lower fragments respectively.

### Orang-utan

Initial sequence analysis of the orang-utan revealed a large number of differences between orL and orU (5.0 and 5.3 kb fragments). PCR primers (OrangUp and OrangLow, Fig. 4) were designed and used in combination with one of the primers TSK8K on the four orang-utan DNAs. The one animal, heterozygous for orL and orU, generated amplification products with both primers. The other three orang-utans only generated PCR products in the presence of the OrangLow primer (data not shown). This suggested that orL and orU contain diverged sequences, and so sequence analysis was carried out on both.

Contiguous sequence data (2846 bp) was obtained from orU (EMBL accession no. Z96952). A novel GC-rich (70%) minisatellite consisting of eight 56 bp repeat units, with sequence variation between the repeats, was identified (Fig. 4). Sequences orthologous to the human telomere-adjacent minisatellite and to the SINE were also present in the orU fragment. However, only two minisatellite repeat units were present, and sequence variation suggests that the central two repeat units are absent in the orU fragment. The orU SINE sequence is 280 bp longer than in humans and extends past the position equivalent to the start of the human telomere (Fig. 4). Immediately distal to the SINE is an array of at least 31 telomere-like repeats.

A total of 2182 bp of sequence data was obtained from orL (EMBL accession no. Z96953). Surprisingly, the orL allele contains the same 45 bp duplication observed in the chimpanzee sequence (Fig. 4). The 56 bp minisatellite repeat sequence detected in the orU allele is present, as an array of 15 repeat units that also showed sequence variation between repeats. Further PCR analysis has shown that the orang-utan-specific minisatellite is variable, with four different length alleles observed among three orang-utans. Sequences similar to the human telomere-adjacent minisatellite are present in orL and, like the orU fragment, the central two repeats are absent. However, the SINE is completely absent in orL and instead an array of telomere-like repeats was identified distal to the minisatellite (Fig. 4). These telomere-like repeats consist of diverged repeats often reiterated into a higher order structure such as (TCCGGGTCCGGGTCAGGG)\(n\), in the 39 repeats sequenced. Amplification with a telomere repeat primer in conjunction with a 5′ radioactively labelled primer in the minisatellite region (PrimA, Fig. 1) revealed a ladder of fragments extending for at least 100 repeat units (data not shown). This confirmed that the smear of products shown in Figure 3 was generated from a long array of telomere-like repeats. Furthermore, preliminary PFGE analysis suggested that the long array of telomere-like repeats is located on the terminal restriction fragment in the orang-utan, indicating that the telomere-like repeats in orL represent a diverged region at the proximal end of the orang-utan Xp/Yp telomere.

Comparison of the orU and orL sequences revealed a large number of differences resulting in 2% divergence (Fig. 4; Table 1). The high level of sequence variation (one every 50 bp) between orU and orL is similar to that observed between haplotypes adjacent to the human Xp/Yp telomere. However, in contrast to the human Xp/Yp haplotypes, there was no significant clustering of sequence polymorphism in the orang-utan (Fig. 4). It is noteworthy that the two divergent orang-utan sequences are probably adjacent to the telomere, possibly suggesting that the same underlying forces have been involved in the evolution of these sequences in humans and orang-utans.

Two alternative explanations for the differences between orL and orU can be considered. Firstly that orL and orU represent a polymorphic duplication of the telomere-adjacent DNA. This seems unlikely because there is no evidence of a duplication of the \( D Y S S 1 4 \)-like locus in orang-utans (see Fig. 2). In addition, no more than two alleles per individuals have been detected at the orang-utan-specific minisatellite but, if the sequence was duplicated, additional DNA fragments might be detected. Secondly, orL and orU could be derived from different orang-utan subspecies. Two separate subspecies of orang-utan exist in Borneo (\( P o n g o ~ p y g m a e u s ~ p y g m a e u s \)) and north Sumatra (\( P o n g o ~ p y g m a e u s ~ a b e l i i \)). The subspecies are characterised on the basis of morphological, behavioural and cytogenetic differences (23,24). This categorisation has been supported recently by molecular data, which suggest the two subspecies diversified 1.5–1.7 million years ago (25). Since the two orang-utan subspecies can form fertile hybrids (26), it remains possible that the heterozygous orang-utan containing orL and orU was the offspring of an inter-subspecies mating. If so, then orL and orU would represent subspecies-specific structures, as for example the pericentric inversion of orang-utan chromosome 2, that has been used for subspecies classification (24). At the time

### Table 1. Pairwise comparisons between DNA sequences orthologous to the 2007 bp of DNA immediately adjacent to the human Xp/Yp telomere

<table>
<thead>
<tr>
<th>Species compared</th>
<th>Sequence compared (bp)</th>
<th>Percentage divergence</th>
<th>Previously published % sequence divergences (difference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>hu/ch</td>
<td>1966</td>
<td>3.5</td>
<td>1.61 (2.2)</td>
</tr>
<tr>
<td>hu/gor</td>
<td>1886</td>
<td>4.0</td>
<td>1.72 (2.3)</td>
</tr>
<tr>
<td>hu/orL</td>
<td>2563</td>
<td>9.6</td>
<td>3.39 (2.8)</td>
</tr>
<tr>
<td>hu/orU</td>
<td>2413</td>
<td>10.4</td>
<td>3.39 (3.1)</td>
</tr>
<tr>
<td>ch/gor</td>
<td>1932</td>
<td>4.2</td>
<td>1.84 (2.3)</td>
</tr>
<tr>
<td>ch/orL</td>
<td>2556</td>
<td>9.4</td>
<td>3.52 (2.7)</td>
</tr>
<tr>
<td>ch/orU</td>
<td>2372</td>
<td>9.8</td>
<td>3.52 (2.8)</td>
</tr>
<tr>
<td>gor/orL</td>
<td>2499</td>
<td>8.8</td>
<td>3.47 (2.5)</td>
</tr>
<tr>
<td>gor/orU</td>
<td>2225</td>
<td>9.2</td>
<td>3.47 (2.7)</td>
</tr>
<tr>
<td>orL/orU</td>
<td>2383</td>
<td>2.0</td>
<td>Average (2.6)</td>
</tr>
</tbody>
</table>

The sequence differences between the species were distributed across the region, and between the two orang-utan structures. Comparisons were made and divergence estimates obtained using Bestfit (Genetics Computer Group version 6.2). Previously published sequence divergence estimates from the \( \varphi \eta \)-globin gene region were included for comparison (5). hu, human haplotype A, ch, chimpanzee, gor, gorilla and orU/orL, orang-utan upper and lower fragments respectively.
that blood samples were collected, the two orang-utan subspecies were not managed separately in captivity, as is now the case, and the subspecies classification of the orang-utans analysed in this study are therefore unknown.

**Sequence comparison and ancestral structures**

Pairwise comparisons were made between the sequences obtained from the higher primates. Higher levels of sequence divergence were observed compared with other loci (Table 1) such that, on average, the between-species divergence is 2.6-fold higher across the subterminal Xp/Yp sequences than at the \( \psi \eta \)-globin pseudogene or other non-coding loci (5,27–29).

Sequence variation was identified in the gorilla and orang-utan DNAs, but none of the variant positions were the same as those identified in humans. In order to derive the ancestral state for the human Xp/Yp telomere-adjacent haplotypes identified at the Xp/Yp telomere junction, the sequences of the polymorphic sites that define the human haplotypes were compared in each of the species examined (Table 2). In all but three positions, there is complete concordance between the species. Therefore, these polymorphic positions have probably been derived since the divergence of the Pan and Homo lineages. At three positions (–1888, –338 and –146, Table 2), the sequence was different between the species. At two of these positions (–338 and –146), the bases found are the same as in the divergent human haplotypes.

**DISCUSSION**

Here we have described the characterisation of sequences orthologous to the human Xp/Yp telomere from the chimpanzee, gorilla and orang-utan genomes. The sequence divergence between species was higher (2.6×) than that identified in other regions of the genome. There is evidence that gross chromosomal rearrangements, including truncations and exchanges between subterminal sequences of Xp/Yp and autosomes, have occurred, resulting in sequence organisation that is different in each of the four lineages analysed. These data show that a variety of mutational processes have brought about many changes in the Xp/Yp telomeric region over a relatively short period of evolutionary time.

**Table 2.** Comparison of the sequences at the positions that define the human Xp/Yp telomere-adjacent haplotypes with those found in chimpanzees, gorillas and orang-utans

<table>
<thead>
<tr>
<th>Polymeric positions identified in humans</th>
<th>Position from first telomere</th>
<th>Repeat in humans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orang-utan lower</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Orang-utan upper</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Caucasian haplotype B</td>
<td>A X</td>
<td>C A</td>
</tr>
</tbody>
</table>

*The 10 bases between -363 and -373 are present (+) or absent (−) as indicated.

nd = sequence not determined; *sequences not present in non-human primates \( \kappa \) position not in complete linkage disequilibrium

**Deriving the ancestral structure**

The overall distance between DXYS14 and a point defined by the start of the human telomere was probably 5.0 kb in the common ancestor. The sequences deleted from the human lineage to reduce this distance to 3.6 kb and the sequence inserted into orU to increase it to 5.3 kb have not been identified but must lie between DXYS14 and the start of the region we have sequenced.

The SINE sequence is truncated or absent in each lineage, with the exception of the orang-utan upper allele (orU). Therefore, the common ancestor probably contained a longer SINE sequence. The SINE has since been truncated at a different site in each lineage by different sequences. The subterminal satellite truncates the SINE more in the chimpanzee than the Xp/Yp telomere does in the human genome and, therefore, it is unlikely that a common ancestor contained an array of subterminal satellite at this chromosome end. No subterminal satellite sequences were identified in the gorilla sequences characterised in this report. However, subterminal satellite may be present in a more distal location because in situ hybridisation suggested that a block of subterminal satellite is present at the end of Xp/Yp in the gorilla (Di and Royle, unpublished observations). In the gorilla, the SINE was truncated still further by an array of telomere-like repeats, distal to which the gorilla sequence was homologous to sequences from the subterminal regions of human chromosomes 4p, 18p and 4q. The human 18p and 4q sequences contain an array of telomere-like repeats proximal to non-repetitive DNA, and the sequence is preserved in the gorilla in the same orientation and position relative to the flanking DNA (20–22). A complex set of events must have occurred in the gorilla lineage to move subterminal sequences, and perhaps a distal array of subterminal satellite, from the end of an ancestral autosome to the end of Xp/Yp. In the orang-utans, two different structures, orL and orU, were identified at the Xp/Yp subterminal region. Immediately distal to the minisatellite sequence, orL contained telomere-like repeats that are probably at the start of the telomere. The SINE is absent from orL, whereas orU contained a SINE sequence, distal to which were some diverged telomere-like repeats, but it is not known whether they also represent the start of an Xp/Yp telomere. As discussed above, we cannot be certain that the orU structure has the same subspecies origin as the orL structure.
Sequences equivalent to the human telomere-adjacent minisatellite were present in all the species analysed. It is monomorphic with four repeat units in humans and chimpanzees but just two repeats in both the orang-utan structures. However, in gorillas, this minisatellite has two alleles of three and four repeats. Minisatellites often exhibit length variation and they can be relatively transient over short periods of evolutionary time (30). Therefore, it is not possible to deduce the ancestral state of this minisatellite with any confidence. Curiously, a 45 bp duplication was identified in the chimpanzee and orang-utan orL but was absent in the human, gorilla and the orang-utan orU structures. The ancestral structure to these lineages could have contained this duplication, which has since been lost independently from three different lineages. Alternatively, the duplication may have occurred independently in the chimpanzee and orang-utan orL lineages.

These data suggest an ancestral structure for the Xp/Yp telomere junction region of the higher primates (Fig. 5). They also suggest that the terminal sequences of the great ape chromosomes are relatively transient, such that subterminal sequence organisation is unique to a lineage. This dynamism may reflect the ability of telomerase to heal broken chromosomes, resulting in the truncation of the chromosome and repositioning of the telomere, as seen for example in the human and possibly the orang-utan orL structure. However, the presence of subterminal satellite and structures other than a telomere in the chimpanzee and the gorilla indicates that the dynamics of subterminal sequences have also influenced the relative position of the Xp/Yp telomere.

The orang-utan-specific minisatellite

A minisatellite composed of a 56 bp repeat was identified in both the orU and orL structures but it is present as only a monomer unit in humans, chimpanzees and gorillas. The transient nature of hypervariable minisatellites has already been described (30), so the appearance of a novel minisatellite only in orang-utans is in keeping with that observation. The minisatellite shows sequence variation between repeats, as well as repeat copy number variation which is probably amenable to analysis using minisatellite variant repeat mapping [MVR-PCR (31,32)]. Although the full extent of the variability at this locus has not been determined, it may be useful for the analysis of orang-utan population structure in both wild and captive populations.

Sequence divergence

Unusually high levels of sequence divergence have been identified in the DNA orthologous to the human Xp/Yp telomere junction. In the human and probably the orang-utan genomes, the sequence is adjacent to a telomere. In contrast, the orthologous sequence is adjacent to a heterochromatic block of subterminal satellite in the chimpanzee genome and probably close to one in the gorilla genome. Little is known about the evolution of single copy sequences from heterochromatic regions, but it is possible that close proximity to heterochromatin or to a telomere has influenced the levels of inter-specific or inter-haplotype sequence divergence (33).

Sequence ancestral to the human Xp/Yp telomere-adjacent haplotypes

Sequence variation was observed in both gorilla and orang-utan; none of the polymorphic sites detected in these species were the same as the polymorphic positions that define the human Xp/Yp haplotypes. The likely ancestral sequence for the human telomere-adjacent haplotypes has been determined by comparison of the Xp/Yp sequence data obtained from the chimpanzee, gorilla and orang-utan genomes (Table 2). It has not been possible to determine the ancestral state at three of the human polymorphic positions. Two of the sites (–338 and –146) are fixed for one or other of the polymorphic bases in the other three species. While it is tempting to speculate that these polymorphisms pre-date the separation of the human from the other lineages, the majority of the polymorphic positions are not shared between species and are therefore more recent. These data indicate that the highly diverged haplotypes in the DNA immediately adjacent to the human Xp/Yp telomere have evolved after the separation of the homo and pan lineages. The mechanisms underlying the generation and maintenance of such highly diverged haplotypes remain unresolved.

MATERIALS AND METHODS

Genomic DNAs

Human DNAs were lymphoblastoid DNAs from 80 unrelated individuals which constitute the parents of the CEPH families. Chimpanzee, gorilla and orang-utan genomic DNAs were extracted from blood and tissue samples obtained from captive bred animals.

PCR

PCRs were carried out using the buffer system described previously (32). Amplifications were carried out using a PTC-200 thermal cycler (MJ research) at 96°C for 20 s, 65°C for 30 s and 70°C for 1–4 min for 32 cycles; the annealing temperature was reduced to 55°C for weak or unsuccessful amplifications. DNA fragments from successful amplifications were resolved by agarose gel electrophoresis and purified from an agarose gel slice by centrifugation through glass wool.
Isolation of distal Xp/Yp sequences.

Isolation of distal Xp/Yp sequences from the gorilla genome was achieved by nested vectorette PCR (19) using the primer PrimB and the vectorette primer, followed by primer TSK8N with the vectorette primer. The vectorette was ligated to a Sau3AI site located 400 bp distal to the start of the interstitial block of telomere repeats identified in gorilla. The vectorette Vec1/MboI primer system has been described previously (34).

Isolation of the chimpanzee sequences distal to the point of divergence from the human sequence was achieved by PCR amplification using a 5′ radioactively labelled primer PrimA [labelled in 10 µl reactions with 0.37 MBq (γ-32P)ATP (DuPont NEI), 5 U of T4 polynucleotide kinase (GIBCO/BRL) and the manufacturers forward labelling buffer system] into the subterminal repeat array using primer Ch7B. The ladder of labelled fragments was resolved in denaturing polyacrylamide gel. 5′ Radioactively labelled single-stranded DNA fragments were isolated from the polyacrylamide gel by removing the polyacrylamide containing the fragment from the backing paper. The polyacrylamide was soaked in 10 µl reactions with 0.37 MBq [labelled in 10 µl reactions with 0.37 MBq (γ-32P)ATP (DuPont NEI), 5 U of T4 polynucleotide kinase (GIBCO/BRL) and the manufacturers forward labelling buffer system] into the subterminal repeat array using primer Ch7B. The ladder of labelled fragments was resolved in denaturing polyacrylamide gel.

The primer sequences used were: TSK8M, 5′-ACACAGTTT-TCCAGTTGTGTT-3′; TSK8N, 5′-TCAAGGAGCTGGTGCA-GAC-3′; TSK8S, 5′-CACAAGAGCCTCCAGGCCAG-3′; TSK8R, 5′-CTTTCCCACTGCTGTTTCCCT-3′; TSK8L, 5′-TCTTT- TTACATGCTAATTCTGAT-3′; PrimA, 5′-gcggtaCCCAA-AGACAGAAGGGGCCAG-3′; PrimB, 5′-gcggtaGAACTGGGGTTATCGACCGAGT-3′; Ch7B, 5′-CCGCTATCTGTTATAAAC-ATGGA-3′; and 29c1B, 5′-TCTCTTTTTGTGCTCCGTCC- CATC-3′. Lower case sequence represents 5′ non-complementary tails containing restriction sites. For other primer sequences see (4).

The allele-specific primers used were: Gor4A, 5′-GTTACA- GATGAGGAGCTCAC-3′; Gor4B, 5′-GTGAGCCTCTCCTCAGT- GAAC-3′; OrangUp, 5′-GCAATCCGCCAGTACG-3′; OrangLow, 5′-ACATCCGGGAATTGCGGA-3′.

DNA sequencing

Purified double-stranded DNA fragments (20–50 ng) were sequenced using Applied Biosystems DNA sequencing platforms and Applied Biosystems dye terminator sequencing chemistries. Contiguous sequences were assembled using the Applied Biosystems autoassembler programme (version 1.4) for Macintosh.

Southern analysis

Genomic DNA was digested to completion with restriction enzymes according to the manufacturer’s recommendations (GIBCO/BRL). The digested DNAs were resolved by agarose gel electrophoresis and blotted onto Hybond-Np membrane (Amersham). Hybridisations were carried out using 32P-labelled DNA probes (random hexaprime labelled) in the `blotto' hybridisation mix (35), at 65°C overnight. Unique sequence DNA probes were hybridised to genomic DNA Southern blots in 5× Denhardt’s, 5× SSC and 1% SDS at 65°C overnight (35).

ACKNOWLEDGEMENTS

We thank A. J. Jeffreys, C. May, P. Bois and J. Buard for stimulating discussions and Y. Dubrova for help with statistics. N.J.R. is a UK-HGMP Senior Research Fellow. This work was funded by the Medical Research Council of Great Britain.

REFERENCES