Sequences flanking the centromere of human chromosome 10 are a complex patchwork of arm-specific sequences, stable duplications and unstable sequences with homologies to telomeric and other centromeric locations

M. S. Jackson1,* , M. Rocchi2 , G. Thompson1 , T. Hearn1 , M. Crosier1 , J. Guy1 , D. Kirk1 , L. Mulligan3 , A. Ricco2 , S. Piccininni2 , R. Marzella2 , L. Viggiano1,2 and N. Archidiacono2

1Department of Human Genetics, University of Newcastle upon Tyne, 19/20 Claremont Place, Newcastle upon Tyne NE2 4AA, UK, 2Istituto di Genetica, Via Amendola 165/A, 70126 Bari, Italy and 3Department of Paediatrics, Queen's University, 20 Barrie Street, Kingston, Ontario K7L 3N6, Canada

Received July 28, 1998; Revised and Accepted October 2, 1998

Little is known about sequence organization close to human centromeres, despite empirical and theoretical data which suggest that it may be unusual. Here we present maps which physically define large sequence duplications flanking the centromeric satellites of human chromosome 10, together with a fluorescence in situ hybridization (FISH) analysis of pericentromeric sequence stability. Our results indicate that the duplications on each chromosome arm are organized into two blocks of ∼250 and 150 kb separated by ∼300 kb of non-duplicated DNA. The larger proximal blocks, containing ZNF11A, ZNF33A and ZNF37A (10p11) and ZNF11B, ZNF33B and ZNF37B (10q11), are inverted. However, the smaller distal blocks, containing D10S141A (10p11) and D10S141B (10q11), are not. A primate FISH analysis indicates that these loci were duplicated before the divergence of orang-utans from other Great Apes, that a cytogenetically cryptic pericentric inversion may have been involved in the formation of the flanking duplications and that they have undergone further rearrangement in other primate species. More surprising is the fact that sequences across the entire pericentromeric region appear to have undergone unprecedented levels of duplication, transposition, inversion and either deletion or sequence divergence in all primate species analysed. Extrapolating our data to the whole genome suggests that a minimum of 50 Mb of DNA in centromere-proximal regions is subject to an elevated level of mechanistically diverse sequence rearrangements compared with the bulk of genomic DNA.

INTRODUCTION

The centromeric regions of most eukaryotic chromosomes are rich in repetitive satellite DNAs (1). Analyses of these regions in humans have focused largely on the alpha satellite present in tandem arrays at the primary constriction of all human chromosomes (2) and its role in centromere function (2,3). Other centromere-specific satellites have been identified in humans (4) and, although these have no known function, some are highly conserved between species (5). However, the physical organization of these satellites has only been established at a few centromeres by pulsed-field gel electrophoresis (PFGE) or fluorescence in situ hybridization (FISH) (6–10). As a result, the boundaries between centromeres and chromosome arms are probably the most poorly characterized regions of the human genome.

Junctions between centromeric satellites and genes are of interest for several reasons. Position effects, most frequently caused by the juxtaposition of constitutive heterochromatin and a transcriptionally active gene (11), have now been identified in mammals including humans (12). It is also becoming apparent that sequence repetition per se can silence transcription in mammalian cells (13,14). Investigating the position and regulation of genes close to centromeric repeats is, therefore, an important step in understanding chromatin compartmentalization.

These regions are also of interest because of the different ways in which satellites and genes evolve. Satellite sequences, such as

*To whom correspondence should be addressed. Tel: +44 191 222 8005; Fax +44 191 222 6662; Email: mjackson@hgmp.mrc.ac.uk
the alpha satellite, undergo concerted evolution where sequence identity is maintained within a species by frequent inter- and intrachromosomal sequence exchanges (15). This can lead to striking differences between the satellites of closely related species. For example, when alpha satellite probes specific for each human chromosome are hybridized to metaphase chromosomes from other primates, a striking lack of positional conservation is observed; all but the alpha satellite probe derived from the human X chromosome hybridize to phylogenetically different chromosomes in chimpanzee and gorilla (16). This is in sharp contrast to the small genetic alterations characteristic of gene evolution (17) and observed; all but the alpha satellite probe derived from the human from other primates, a striking lack of positional conservation is maintained between the chromosome arms of distantly related species (18). Therefore, it is of interest to determine how, and to what extent, the mutational processes associated with satellite sequence evolution are partitioned from the surrounding DNA at human centromeres.

There is already some evidence that sequences close to centromeric satellites may be unusual. Extensive duplications flanking centromeric regions have been observed directly or inferred on human chromosomes 1, 2, 9 and 10 (19–24). More recently, tracts of gene-related sequences have been characterized which have undergone repeated transposition events into or between pericentromeric locations in the human genome. These include sequences related to the ALD, NF1, KGF and creatine transporter genes (25–28). This pattern of evolution, together with the fact that the pericentromeric NF1 pseudogenes are present only on chromosomes which contain alpha satellite from superchromosomal family 2 (27,29), suggests a possible link between these duplications and satellite repeats. However, the lack of accurate maps of these regions precludes any integrated analysis of these phenomena.

Recently, we published a 9.75 Mb PFGE and yeast artificial chromosome (YAC)-based map across the centromere of human chromosome 10 (9). This linked duplicated ZNF gene clusters sharing ~96% sequence identity to centromeric satellites on both chromosome arms. The duplicated loci ZNF11A, ZNF33A and ZNF37A map to 10p11, while ZNF11B, ZNF33B and ZNF37B map to 10q11–11.2 (19,30). In addition, duplicated copies of the marker D10S141 were shown to map distal to the ZNF clusters on both arms (9). This map provides an opportunity to investigate sequence organization and evolution across the entire pericentromeric region of a human chromosome. In the present report, we establish the size and orientation of the sequence duplications flanking this centromere in human, the proximity of these duplications to repetitive sequences and the stability of these sequences in other primates. Our results indicate that the pericentromeric region of chromosome 10 consists of a complex patchwork of chromosome arm-specific sequences interspersed with 100–250 kb tracts of stable duplicated DNA and highly unstable sequences with homology to other genomic locations. Analyses of other primates suggests that the entire region is prone to rearrangement, with evidence for transposition, inversion and either the deletion or rapid divergence of sequences in all lineages analysed.

RESULTS

The ZNF and D10S141 duplications are ~240 and 150 kb in size, respectively

To estimate the extent of the ZNF and D10S141 duplications on chromosome 10, we analysed the ZNFB and D10S141B loci present within the q arm YAC contig by PFGE. This contig contains overlapping non-chimeric clones with internally consistent restriction maps (9,31). The duplicated or unduplicated nature of each marker was first established by hybridization to a YAC panel which spans the p arm (A) and q arm (B) copies of each duplication (Table 1; ref. 9). Markers were then placed on the existing PFGE map as described previously (9). To confirm that unique markers at the termini of duplicated sequence do not represent small insertion or deletion events, clones suitable for FISH were isolated (see Materials and Methods) and hybridized to human metaphase spreads to establish if they identified sequences on one or both sides of the centromere. The YAC map from 10q11 spanning the ZNFB and D10S141B duplications is presented in Figure 1A, while the hybridization and FISH data which define the extent of duplicated sequence flanking the centromere are presented in Figure 1B and C, respectively.

The proximal boundary of the ZNFB duplication lies between 274λAT3B which is duplicated (Fig. 1B, panel 2) and 236A11L which is not (Fig. 1B, panel 1; and C, panel 1). The distal boundary maps between 214H10T and WME31-R (Fig. 1A). Although 214H10T is unique within the YAC contig (Fig. 1B, panel 6), a phage clone positive for 214H10T hybridizes to both sides of the centromere (Fig. 1C, panel 3), indicating that it contains duplicated sequences. However, WME31-R which maps 5–20 kb distal of 214H10T (Fig. 1A) only identifies sequences in 10q11 by FISH (Fig. 1C, panel 4). The estimated physical size of the ZNFB duplication on chromosome 10 is, therefore, 230 ± 20 kb (the distance between the 274λAT3/236A11L interval and the 214H10T phage clone).

The proximal boundary of the D10S141B duplication lies between 313F4PB which is duplicated on chromosome 10 (Fig. 1B, panel 8) and 918F7R which is not (Fig. 1B, panel 7; and C, panel 5). However, the 918F7R FISH probe also hybridizes to the centromeric regions of three other human chromosomes (1, 2 and Y; see later) in addition to chromosome 10, indicating that the proximal boundary of the D10S141B duplication currently is defined by sequences which are not unique within the genome. The distal boundary of the D10S141B duplication maps within the cosmid RT3-2 as a fragment from the proximal end of this cosmid is duplicated (RT3-2.H/X5B; Fig. 1B, panel 12) while more distal fragments are not (Fig. 1B, panel 13; and C, panel 7). Thus, the estimated size of the D10S141B duplication is between 110 kb (the distance between 313F4P and RT3-2.H/X5B) and 180 kb (the distance between 918F7R and RT3-2.E8).

The ZNF and D10S141 duplications are not in the same orientation

We have placed a total of 36 markers on the existing PFGE map of the region (9) to establish the relative position of the duplicated loci in 10p11–q11 (see Materials and Methods). These data are presented in Figure 2. The ZNFA and ZNFB duplications are known to be inverted relative to each other (9; Fig. 2, blue arrows). The position of additional markers from these duplications (237C10L, D10S1746, 746H6b/B4, 274C3/4T3 and 274C4X7T3) confirms this orientation. Importantly, our new markers establish the orientation of the D10S141A and D10S141B duplications. The p arm markers 313F4P and CM5F9A map distal to 367c4.4A as they are present on 250 and 50 kb NruI fragments, respectively, in YACs y837B5, y985D10 and y746H6 (Fig. 2). The markers 367E5A and RT3-2.H/X5A map to the same NruI fragment as 367c4.4A in YACs...
Figure 1. (A) Physical map of the ZNF6 and D10S141B duplicated sequences. The positions of clones used in the construction of the map are shown as solid horizontal bars at the top. Markers are placed between the enzyme sites and/or YAC end clones which define their position. ZNF37B and D10S1746 map to the same interval indicated by a solid line joining these markers. Enzymes used are as follows: C, Clal; K, KpnI; M, MluI; F, SfiI; S, SalI; Sc, SacI; Sm, SmaI; V, EcoRV; X, XhoI. An MluI site identified by partial digestion is indicated in parentheses. Not all restriction sites are shown, with the exception of MluI. Due to the low density of markers around D10S141B, a cosmid contig was constructed using clones previously isolated from the region (62). This contig was restriction mapped using EcoRI, HindIII and BamHI (data not shown) and integrated with the YAC-derived PFGE map using the SfiI sites which are present within both the YAC and cosmid contigs. Gel-isolated restriction fragments from cosmids provided additional markers (see Materials and Methods) the position of which are shown by vertical dotted lines. The positions of cloned YAC ends (L = left, R = right) are indicated by dashed arrows. The minimal extent of the duplications is indicated by dark grey boxes, the maximal extent by light grey boxes. (B) Hybridizations defining the extent of the ZNF6 and D10S141B duplication. The probes used for each hybridization are indicated by the dotted lines joining markers in (A) to individual panels. Lanes p and q denote Southern hybridizations of probes to YACs on the p and q arms (see ref. 9 for full YAC panels). The YACs shown here are: p-y746H6 (all lanes), q-y918F7 (lanes 1–6) q-y214H10 (lanes 7–13). The approximate size of hybridizing fragments is given in kilobases in each case. (C) FISH hybridizations defining the ZNF6 flanking duplications. Individual DAPI-banded chromosomes 10 taken from human metaphases (not shown). All probes are phage clones (insert size 17–22 kb) which encompass the indicated marker, with the exception of cM9 which is a cosmid clone with a 37 kb insert (A and Materials and Methods). Probes are as follows: 1, 236A11L (10q11 signal); 2, ZNF33B (10p11 and 10q11 signals); 3, 214H10T (10p11 and 10q11 signals); 4, WME31-R (10q11 signal); 5, 918F7R (10q11 signal); 6, cM9 (10p11, 10q11 signal); 7, RT3-2.E8 (10q11 signal).

y985D10 and y746H6 but are not present in YAC y837B5. The order of all markers within the D10S141 duplications is therefore, 10pter–313F4PA–cM5.H9A–367c4.4A–[367.E5A/RT3-2.H/X5A]–10cen–313F4PB–cM5.H9B–367c4.4B–367.E5B–RT3-2.H/X5B–10qter. Thus, the p and q arm D10S141 sequences are in the same orientation relative to each other (Fig. 2, red arrows), in contrast to the ZNF A and B duplicated sequences. It is also clear from Figure 2 that the ZNFPA and D10S141A duplications are separated by ~250–320 kb of non-duplicated DNA (from RT3-2.H/X5A to ZNF11A/33A) and that a similar physical interval separates the D10S141B and ZNFB duplications (from 214H10T to the 918F7R-313F4PB interval). It has been proposed that the ZNF and D10S141 loci were duplicated in one event and subsequently inverted (19). However, these results indicate that additional, or more complex, rearrangements must be invoked to account for the current organization of these loci.

FISH analysis of other primates

To investigate the nature and timing of these rearrangements, we have performed a comparative FISH analysis of other primates. In the light of recent reports of transposition between pericentromeric regions (25–28), and the proximity of the duplications to classical satellite sequences (Fig. 2), we developed probes from across the entire pericentromeric region to investigate sequence stability (see Materials and Methods). A total of 22 probes spaced over a 1.5 Mb region of each chromosome arm (Fig. 2) were...
analysed in seven primate species. The results are summarized in Table 1. Partial metaphases from some of these hybridizations are also shown in Figure 3. In the Great Apes, chromosomes are defined by their phylogenetic number (roman numerals); the phylogenetic X chromosome corresponds to human chromosome 10 (HSA10).

Looking at the FISH results as a whole, only four probes out of a total of 22 give comparable hybridization signals in all of the primates analysed (A4C5L, ZNF25, ZNF37A and 746B/B.4). The other 18 probes all provide evidence for sequence rearrangements in one or more primate species. These include inversions (implied by changes in marker order relative to the centromere), duplications and transpositions (implied by hybridization to new or multiple locations in different species), and deletions or rapid sequence divergence (implied by a complete absence of hybridization signal in some species). In general, there appears to be a reduction in sequence stability close to centromeric satellites, although probes up to 1.5 Mb from satellite 3 sequences show some evidence of rearrangement. The high frequency of such mechanistically diverse rearrangements makes it impossible to determine the precise evolutionary history of the ZNF and D10S141 loci.

**Markers have changed position relative to the centromere during primate evolution**

The precise linear order of our probes is only known in human. However, their position relative to the primary constriction at the centromere is clear for all species analysed (Table 1). In Pan troglodytes (PTR), Gorilla gorilla (GGO) and Pongo pygmaeus (PPY), the probes A4C5L, ZNF25, ZNF37A and 746B/B.4 all hybridize to regions homologous to HSA10p11 (Xp11 in HSA, PTR and PPY, Xq11 in GGO). This contrasts sharply with the results for the three Old World monkey species [Macaca fascicularis (MFA), Macaca mulatta (MMU) and Presbytis cristata (PCR)] where one or more of these probes hybridize specifically to regions homologous to HSA10q11. In PCR, the centromere can be mapped to between D10S141A and A4C5L (Table 1). In MFA and MMU, the centromere is in a similar location but, as the origin of the single hybridization signal obtained with D10S141A is unknown in this species (it could be from the paralogue of D10S141A, D10S141B or from both), it is not possible to map the centromere accurately relative to this marker (dotted lines, Table 1). Among the Great Apes, one marker, 331F5L, has also changed position relative to the centromere. In Homo sapiens (HSA), PTR and PPY, the centromere maps between 906Y1.29 and 331F5L whereas in GGO it maps between 331F5L and 236A11L.

**The ZNF and D10S141 duplications occurred before the divergence of orang-utan from other Great Apes**

Five probes are derived from loci within the human duplications (bold in Table 1, row 1). These give two signals which flank the centromere of the phylogenetic X chromosome in most hybridizations to Great Apes (Fig. 3A). This indicates that the ZNF and D10S141 sequences were duplicated before the divergence of orang-utan from the other Great Apes. A notable exception is the ZNF37A probe which only hybridizes to Xp11 in humans despite...
Table 1. Comparative FISH analyses of chromosome 10 pericentromeric probes

<table>
<thead>
<tr>
<th>Probe</th>
<th>HSA</th>
<th>PTR</th>
<th>GGO</th>
<th>PPy</th>
<th>MFA</th>
<th>MMU</th>
<th>PCr</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yp11</td>
<td>Yp11</td>
<td>Yp11</td>
<td>Yq11</td>
<td>Ypl</td>
<td>Yq11</td>
<td>Yq11</td>
</tr>
<tr>
<td>5</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
</tr>
<tr>
<td>10</td>
<td>Yp11</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
</tr>
<tr>
<td>15</td>
<td>Yp11</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
</tr>
<tr>
<td>19</td>
<td>Yp11</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
</tr>
</tbody>
</table>

HSA, Homo sapiens (human); PTR, Pan troglodytes (chimpanzee); GGO, Gorilla gorilla (gorilla); PPy, Pongo pygmaeus (orangutan); MFA, Macaca fascicularis (crab-eating macaque); MMU, Macaca mulatta (rhesus monkey); PCr, Presbytis cristata (silvered leaf monkey).

In the apes, chromosomes are defined by their phylogenetic number (roman numerals); human chromosome 10 corresponds to the phylogenetic X chromosome. The nomenclature of cytogenetic bands for the Great Apes is from ISCN 1985. The probes are arranged according to their physical position in human. Probes from duplicated regions are highlighted in bold. Probe 746B/B4 was isolated using a duplicated marker of the same name (Fig. 2 and Materials and Methods) but is classified here as non-duplicated as it identifies sequences in regions homologous to HSA10p11 by FISH. Hybridization of probes from 10p11 to cognate loci in human (Xp11) and to syntenic regions in other species is shown in red, hybridization to duplicated Xq11 sequences is shown in dark blue. Hybridizations of probes from 10q11 to cognate loci in human (Xq11) and syntenic regions in other species are shown in dark blue, hybridizations to duplicated Xp11 sequences are shown in red. Hybridizations to other pericentromeric locations are shown in light blue. Hybridizations to telomeric locations are shown in green. The absence of a hybridization signal is indicated by N.S. (no signal, see Materials and Methods). A dash indicates that no hybridization has been performed.

1GGXq11 is homologous to HSA10p11, GGOXp11 is homologous to HSA10q11 (M.Rocchi, unpublished data).
2MFA10 and MMU10 are homologous to HSA10 (68).
3PCR12p11 is homologous to HSA10p11; PCR12q11 is homologous to HSA10q11 (69; M.Rocchi, unpublished data).

(a) Iqtel double signal, IIptel, IXqtel, XIptel.
(b) IIptel, Vqtel, IXqtel, XIptel.
(c) Double signals on chromosomes II and IX (IIp11, q11; IXp11, q13).
(d) IIptel double signal, IIptel, Vqtel, IXqtel, XIptel.
(e) Iqtel, IIptel, Iqtel, IIptel, Vqtel, VIIptel, XIptel, XVIIptel, Xqtel and other telomeres.
(f) MFA1 is homologous to HSA1.
(g) MMU2, MFA2 and PCR4 are homologous to HSA7.
(h) MMU15 is homologous to HSA2p.
(i) MMU2p is homologous to HSA7 and HSA21: hybridization is to the HSA21 homologous region.
(j) PCR6 is homologous to HSA1 and HSA19: hybridization is to the HSA1 homologous region.
(k) Marked chromosome. The region showing hybridization is homologous to HSA21 (70).

The fact that this gene is known to be duplicated and present on both sides of the centromere (9, 19). This could be due to tracts of unique sequences within this probe or to a +/- polymorphism for ZNF37B which contains multiple frameshifts and stop codons and is presumed to be a pseudogene (T.Hearn, unpublished data).

In the Old World monkey species, all five probes derived from the human duplications hybridize specifically to a single location syntenic to HSA10q11 (10q11 in MFA and MMU; 12p11 in PCR), indicating that sequences related to these probes are confined to a single chromosomal location in these species. One exception is the D10S141A probe which gives two signals flanking the centromere in PCR, a result which implies that further rearrangement of these sequences has occurred. Furthermore, one probe which is specific to the q arm in human, chimpanzee and orangutan (WME31-R) hybridizes to both sides of the centromere in gorilla.

Probes close to the centromere provide evidence for extensive sequence reorganization

Only one probe derived from the human duplications gives a signal on a chromosome not syntenic to HSA10; D10S141A hybridizes to Yq in addition to Xp11 in PPY. This is in sharp contrast to other probes within 400 kb of centromeric satellites which give highly differentiated patterns of hybridization in Great Apes. Two probes from 10q11 in human, 331F5L and 918F7R, hybridize to between one and seven additional pericentromeric locations in all primates analysed (Fig. 3B). The 918F7R probe also maps to Ypter in PTR and to the long arm of the Y chromosome in PPY (Table 1). A total of six probes from the 400 kb interval between ZNFA and satellite 3 sequences in 10p11 (Fig. 2 and Table 1) hybridize to centromeric and subtelomeric locations in most species examined. For example, 746Y1.20 and
746Y1.27 hybridize to a probe-specific combination of centromeric and subtelomeric locations in the Great Apes (Fig. 3C) but map to a single pericentromeric location in MFA and MMU. The more distal probes 746B/B4, 746Y1.6 and 746Y1.32 give a more restricted pattern, frequently hybridizing to chromosome I in Great Apes (Fig. 3D) and hybridizing to a single pericentromeric or subtelomeric location in Old World monkeys (Table 1).

A total of 10 probes failed to give any hybridization signals in one or more primate species (Table 1, N.S.). To confirm that these results were not due to technical failure, the quality of each labelled probe which gave a negative result was checked on human metaphases and then co-hybridized with a control probe back to the metaphases under study (see Materials and Methods). In each case, only signal from the control probe was observed. This strongly suggests that sequences related to these probes are not present in the genomes of the species in question. It is noteworthy that seven of the 10 probes which do not hybridize to Old World monkey chromosomes are probes which hybridize to multiple locations in the Great Apes. However, three probes which do not hybridize to some species, 295SC3, ZNF37A and 236A11L, give relatively stable hybridization patterns in Great Apes, illustrating that the correlation between these hybridization patterns is not perfect.

Although sequence instability is most pronounced close to centromeric satellites, it is not confined to these regions. Three probes which map between the ZNFB and D10S141B duplications in human (WME31-R, 273E3P and 738F9R) all hybridize to a region syntenic to the middle of HSA10p in at least two species (Fig. 3E). For example, the 738F9R probe hybridizes to 10p13 in human (Xp13) and to the homologous region in GGO (Xq24), MMU (10p) and PCR (12p). This suggests that a region of homology between HSA10q11 and HSA10p13 has existed since before the divergence of Old World monkeys and apes. Furthermore, the most distal p arm probe in humans (295SC3) hybridizes to 12pter as well as 12q11 in PCR (Fig. 3F), and 738F9R gives two distinct q arm hybridization signals in MMU implying local sequence duplication or inversion in this species.

**DISCUSSION**

We have shown that a total of 315–440 kb of duplicated sequences flank the centromeric satellites of human chromosome 10. Two duplicated blocks of ~250 kb (containing the ZNFA and ZNFB loci) and 150 kb (containing D10S141A and B) lie on each chromosome arm and are separated by 250–320 kb of non-duplicated DNA in both cases. Our FISH results indicate that these sequences were present on both sides of the phylogenetic X (i.e. HSA10) centromere before the divergence of orang-utan from the other Great Apes ~13–18 millions of years ago. This is consistent with the observed level of sequence divergence between the duplicated genes (19,32,33). The primate FISH also identifies changes in marker order relative to the centromere between Old World monkey species and the Great Apes. Although centromere movement could be invoked to account for these results (3), a simpler interpretation is that a pericentric inversion has occurred after the divergence of these lineages, with one breakpoint close to D10S141A and the other within the centromere (Table 1). Although this inversion could have occurred in either the Old World monkey lineage or the lineage leading to Great Apes, its position and timing would be consistent with a role in the formation of the flanking duplications.

The original model proposed for the evolution of the ZNF and D10S141 loci involved tandem duplication followed by pericentric inversion to place related loci on both sides of the centromere (19). However, the frequency and extent of rearrangements implied by our analyses make it impossible to establish the nature, or timing, of these events. For instance, the duplicate signals obtained with the D10S141A and WME-31 probes in PCR and GGO, together with the fact that some probes duplicated in human (e.g. ZNF37A) only hybridize to one chromosome arm in some Great Apes, indicate that additional rearrangements have occurred. This conclusion is also supported by the fact that 331F5L maps to a different position relative to the centromere in gorilla compared with the other Great Apes, suggesting that a further pericentric inversion has occurred in this species. Such frequent rearrangements make the assumption that the ZNF and D10S141 loci were duplicated in one event (19) questionable. Furthermore, the assumption that sequence duplication would pre-date or coincide with sequence inversion during the evolution of...
of this region (19) is also no longer secure. It recently has been suggested that inversions per se can lead to cytogenetically cryptic duplications and deficiencies as a consequence of aberrant meiotic exchange in inversion heterozygotes (34). Therefore, it is possible that sequence inversions may have occurred before, and may be mechanistically related to, the observed duplication events. It is clear that extensive sequence data, together with comparative analyses at a resolution not currently available for megabases of DNA, will be required to understand the complex rearrangements we have identified.

Although this is the first integrated analysis of pericentromeric sequence organization, there is increasing evidence that duplications and inversions occur frequently in these regions. Some of the most common deletion syndromes in man affect centromere and inversions occur frequently in these regions. Some of the sequence organization, there is increasing evidence that duplications rearrangements we have identified.

Pericentric inversion is also believed to have moved the W region of the human Vk immunoglobulin locus from one side of the centromere of chromosome 2 to the other in the common ancestor of human and chimpanzee (21). Cytogenetic banding patterns in Great Apes are consistent with this hypothesis (38). On chromosome 1, three genes of the FCGR1 family flank the centromeric heterochromatin at 1p12 and 1q21 and share ∼98% sequence identity (23). A pericentric inversion flanking the centromere of this chromosome had also been inferred purely from cytogenetic analysis (38). In addition, cosmids containing the phosphoglucomutase gene PGM5 (22) and the freuc-5 gene (39) both identify sequences flanking the centric heterochromatin on chromosome 9. It is also worth noting that several probes used in the present study hybridize to both sides of some primate centromeres (331F5L, 746Y1.20 and 746Y1.27). Finally, FISH data suggest that ChAB4 and KGF-related sequences flank the centromeres of chromosomes 18 and 21 (28,40).

It is possible that the inversions identified or implied by these analyses are mechanistically related to inversions which are known to occur at high frequency within the pericentromeric heterochromatin of chromosomes 1, 9 and 16 (41,42). Molecular cytogenetic analyses of human chromosome 9 heteromorphisms indicate that the breakpoints of these inversions can occur within a variety of centromeric satellites and that there is a high de novo inversion frequency (43). There is also direct sequence evidence that cytogenetically cryptic inversions involving the alpha satellite have occurred close to the centromere of chromosome 7 (44). It is possible, therefore, that inversions within or close to pericentromeric heterochromatin occur at high frequency at all human centromeres but that most remain undetected due to the low resolution of standard cytogenetic techniques. If this is so, it is plausible that the breakpoints of a small percentage of these inversions occur outside the centromeric satellites, disrupting gene-containing sequences. Since recombination is greatly reduced close to human centromeres (9,45), we would expect potentially deleterious duplications and deficiencies produced by recombination in inversion heterozygotes to be rare. As a result, these inversions would have a high probability of being selectively neutral.

Regardless of the precise events affecting the ZNF and D10S141 sequences, the comparative FISH analysis indicates that they have remained relatively stable components of the chromosome in great ape species, in sharp contrast to other sequences in the region. A summary map showing human sequence relationships is presented in Figure 4. Stable arm-specific sequences are interspersed with large tracts of duplicated DNA and with sequences which exhibit species-specific homologies to other centromeric or telomeric locations. The transience of these homologies suggests that rapid sequence evolution/reorganization is occurring continually in all primate lineages. This reorganization predominately, but not exclusively, involves sequences close to human centromeric repeats. Although the nature of these sequences currently is unknown, the behaviour of the q arm probes 331F5L and 918F7R is typical of low copy number complex repeat sequences which have been identified close to several human centromeres. Examples of these include the ChAb4 family (40,46), the NFI pseudogenes (27), the ALD pseudogenes (26) and the KGF gene segments (28). The predominantly pericentromeric location of these sequences has led to the suggestion that their mobility may be associated intimately with the concerted evolution of centromeric satellites (27). The marker 331F5L is consistent with this hypothesis as it maps between satellite 2 and satellite 3 sequences and hybridizes to seven out of the 15 human centromeres where satellites 2 and 3 have been mapped by FISH (47). The 918F7R probe also hybridizes to four human centromeric regions (1, 2, 10 and Y) which contain 5 bp satellite sequences (6,47). Although this probe maps ~600 kb distal of satellite sequences on 10q, it is possible that its current position may be due to the rearrangements we must invoke to account for the position and orientation of the duplications on this chromosome arm.

Proximity to pericentromeric satellites and/or local sequence rearrangement cannot, however, account for the telomeric sequence homologies exhibited by probes within 400 kb of satellite 3 sequences in 10p11 (Table 1 and Fig. 4). These homologies could define the position of a chromosome fusion event during the evolution of HSA10. Human genes from 10p11–q11 map to three different mouse chromosomes (6,14 and 18) and to two chromosomes within the genome of the prosimian Eulemur macaco (48), indicating that this is a possibility. However, the number of homologies which have now been identified between subtelomeric and pericentromeric regions suggests that extensive sequence exchange can occur without cytogenetic rearrangement. For instance, recently duplicated paralogues of the ALD gene, which maps to Xq28, are present in several pericentromeric locations in human including 10p11 (26; represented by a green box in Fig. 4). Telomere-related sequences have been identified close to the centromere of chromosome 12 in a clone which contains sequences duplicated elsewhere on this chromosome (49). Sequences related to the minisatellite λMS29, which maps to 6p25–pter, have been identified close to the centromere of chromosome 16 (50). In addition, a 9 kb repetitive sequence has been identified on the human acrocentric short arms which is homologous to chimpanzee subtelomeric repeats (51), and FISH data suggest that members of the olfactory receptor gene family, which are clustered predominantly in subtelomeric regions, are also present close to the centromeres of human chromosomes 2, 3, 16, 21 and 22 (52). The presence of telomeric sequence close to human centromeres is also indicated by the observation that a number of YACs containing telomeric repeats map by FISH to the pericentromeric regions of specific human chromosomes including chromosomes 2, 6, 7, 9 and 16 (53). It is likely that at least some of these homologies are due to sequence
Figure 4. Summary of sequence organization across the centromere of human chromosome 10. This map is based on the results presented in Figure 2 and Table 1. The approximate positions of FISH probes are shown, together with the homologies to other human genomic locations. The ALD parologue (26) is not integrated into the map as its precise position is unknown. The sizes of the duplications are to scale and their relative orientations are indicated by solid arrows. The size of the domains of homology currently are unknown and therefore are arbitrary. The scale of the central satellite arrays is reduced for ease of presentation.

Movement or exchange between rapidly evolving pericentromeric sequences (25–28) and subtelomeric sequences which share some organizational features and evolutionary dynamics (54–56).

Characterization of sequence exchanges within centromeric and subtelomeric regions has led to speculation that they could result in the formation of novel genes of evolutionary significance (26,54). While it is true that the chromosome 10 ZNF duplication has led to the formation of at least one novel transcribed sequence (19), gene-related sequences in other duplications and pericentromeric-directed transpositions appear to be pseudogenes (23,25–27). It is interesting, in this respect, that 10 of our FISH probes fail to give any hybridization signal in one or more primate species and that these probes tend to be close to centromeric satellites. *De novo* creation of tens of kilobases of DNA seems an unlikely explanation for this observation, particular in the case of probes such as ZNF37A, which contains gene-related sequences, and 746B/B.4 and 236A11L which exhibit consistent and simple hybridization patterns in most primates analysed. Therefore, it is more plausible to suggest that sequences within these probes have been deleted from the genomes of some primates, or have diverged so rapidly in primary sequence that they do not produce a hybridization signal. This implies that the evolution of pericentromeric (and subtelomeric) sequences not only involves rapid sequence formation and dispersal by duplication and transposition, but also extensive sequence deletion/divergence over a relatively modest evolutionary timescale. This is consistent with the primary sequence in these regions being of little or no significance to the organism. However, this does not necessarily mean that they have no function; the recent identification of highly conserved domains within complex subtelomeric repeats (55–57) raises the possibility that similar, sequence-independent, functional domains await identification at human centromeres.

Although this is the first integrated analysis of pericentromeric sequence organization and stability, a recent bacterial artificial chromosome (BAC) map spanning 21q11 suggests that similar sequence organization exists close to the centromere of other human chromosomes. The centromeric end of this BAC map includes a 200 kb sequence block duplicated in 21q22.1 (58) which is flanked by loci mapping to multiple chromosomes by PCR or to other centromeric locations by FISH (59). The similarity between the 21q11 contig and the map of 10p11–q11, together with the evidence of duplications (20–24) and instability in other pericentromeric regions (26–28), suggests that all human centromere-proximal regions may share similar organizational features. If we assume that the physical scale of the reorganization we have characterized here is typical of all human chromosomes, it means that a minimum of 50 Mb of the human genome (1 Mb for every chromosome arm) is subject to an elevated frequency of gross sequence rearrangements. Establishing the underlying mechanisms of these diverse rearrangements will be difficult because of their scale and complexity. However, an achievable goal will be the development of extensive sequence data from a number of human centromeric regions, despite the difficulty of establishing the chromosomal origin of repetitive genomic clones (54,59). These sequence data will provide a logical starting point for investigations of chromatin structure and transcriptional activity in these regions and should facilitate investigation of the functional relevance, if any, of their unusual sequence organization and evolution.

**MATERIALS AND METHODS**

**Markers**

The following markers were placed on the existing PFGE map of this region (9): 313F4P and 214H10T (31); 274C11L, A4C5L, 237C10L, 331F5L, D10S145, 27BB8R, D10S130 and 236A11L (9); WME31-R (60). In addition, the following probes were
Electrophoresis, Southern transfer and hybridization

All procedures were carried out using standard methodologies (63). PFGE was performed using the CHEF DRII and DRIII systems (Bio-Rad, Hercules, CA).

The conditions and the size ranges resolved were as follows: 5–75 kb: 1% agarose, 0.5x TBE buffer with a pulse time of 0.1–6 s ramped over 1 h at 200 V; 5–450 kb: 1% agarose, 0.5x TBE buffer with a pulse time of 0.9–29 s ramped over 20 h at 200 V.

FISH probe development

To isolate genomic clones suitable for FISH, three YACs which form part of the internally consistent YAC pulsed field map of the region and which span the duplications in 10p11–q11.2 (746H6, 918F7 and 738F9; ref. 9) were shotgun cloned into phage libraries from YAC clones isolated from y746H6 (63) and y906G7 (918F7R and 738F9R); see below); 746B/B.4 (BamHI–BstEII fragment subcloned from y746H6; see below): 367.E5, cM5.H9, RT3-2.E7 and RT3-2.E8 (individual HindIII or EcoRI fragments gel isolated from cosmid clones cMEN367 (D10S141B), cM5 and RT3-2 (61, 62); the nomenclature indicates both enzyme and fragment number, for example cM5.H9 is the fifth largest HindIII–XhoI fragment from RT3-2); 295SC3 and D10S1746 (sequence-tagged sites genetically mapped to the pericentromeric region of chromosome 10; MIT/Whitehead Institute).

Where a locus was shown by Southern hybridization to be duplicated in our YAC panel (9), the nomenclature of Tunnacliffe et al. (19) was employed: the p arm locus being termed A, the q arm locus being termed B.

where developed: 274A4T3 and 274A7T3 [vectorette PCR products obtained from phage clones positive for marker 274C11L using the enzyme HindIII in combination with the bacteriophage T3 primer as described previously (9)]; 746B23 and 476A31 (phage λ clones isolated from y746H6; see FISH probe development): 746Y1.32, 746Y1.27, 746H6Y1.6 and 906Y1.29 (Bstu1Y1 fragments isolated from y746H6 and y906G7; see below); 746B/B.4 (BamHI–BstEII fragment subcloned from y746H6; see below): 367.E5, cM5.H9, RT3-2.E7 and RT3-2.E8 (individual HindIII or EcoRI fragments gel isolated from cosmid clones cMEN367 (D10S141B), cM5 and RT3-2 (61, 62); the nomenclature indicates both enzyme and fragment number, for example cM5.H9 is the ninth largest HindIII fragment from cosmids cM5); RT3-2.H/X5 (the fifth largest HindIII–XhoI fragment from RT3-2); 295SC3 and D10S1746 (sequence-tagged sites genetically mapped to the pericentromeric region of chromosome 10; MIT/Whitehead Institute).

Where a locus was shown by Southern hybridization to be duplicated in our YAC panel (9), the nomenclature of Tunnacliffe et al. (19) was employed: the p arm locus being termed A, the q arm locus being termed B.

The 746A23 and 746A31 phage clones contain novel ZNF-related sequences and were isolated from the y746H6 λFix library by screening with the ZNF cDNAs ZNF11A, ZNF27A and ZNF33B (30) under low stringency conditions. In addition, three previously available cosmid clones were used for FISH: RET (cRET9), cM3 (D10S141B; Kwok, 1994, ref. 62) and 25D101F1 (D10S141B; Kwok, 1994, ref. 62). Restriction mapping and hybridization experiments indicate that cM3 and cM9 consist entirely of sequences duplicated for FISH are chimeric.

The 746A23 and 746A31 phage clones contain novel ZNF-related sequences and were isolated from the y746H6 λFix library by screening with the ZNF cDNAs ZNF11A, ZNF27A and ZNF33B (30) under low stringency conditions. In addition, three previously available cosmid clones were used for FISH: RET (cRET9), cM3 (D10S141B; Kwok, 1994, ref. 62) and 25D101F1 (D10S141B; Kwok, 1994, ref. 62). Restriction mapping and hybridization experiments indicate that cM3 and cM9 consist entirely of sequences duplicated for FISH are chimeric.

Isolation of markers between ZNFA and satellite 3

Sequences between ZNFA and satellite 3 DNA are present on a 340 kb MluI fragment in y746H6 and a 300 kb NotI–MluI fragment in y906G7. (9; and M. S. Jackson, unpublished data). Six 100 µl PFGE blocks of each YAC were digested to completion with the appropriate enzyme(s) and electrophoresed through a 1% LMP PFGE gel using standard conditions. The 340 and 300 kb fragments were excised, washed in T.E. (3 x 30 min), melted at 65°C for 15 min and digested with Bstu1Y1 in 1x Universal buffer (Stratagene). After heat inactivating the enzyme for 20 min at 68°C, the samples were digested with β-agarase for 2 h in 0.5x Universal buffer, extracted twice with phenol–chloroform, precipitated in 2 vol of ethanol on dry ice and resuspended in 10 µl of T.E. After resuspension, 3 µl of each digest was ligated to BamHI pZERO-2 (Invitrogen, Carlsbad, CA), transformed, and screened according to the manufacturer’s instructions. Clones larger than 150 bp were identified by a cracking miniprep (65) and sequenced (see below). Clones with no homology to interspersed repeats were identified using the program Repeatmasker (A.F.A Smith and P. Green, unpublished data) and used for mapping.

Fluorescence in situ hybridization

Metaphase spreads have been obtained from lymphoblastoid or fibroblast cell lines from the primate species listed in Table 1. Chromosome preparations were hybridized in situ with probes labelled with biotin by nick translation, essentially as described (66), with minor modifications (67). Briefly, 200 ng of labelled probe were used for each experiment; hybridization was performed at 37°C in 2x SSC, 50% (v/v) formamide, 10% (w/v) dextran sulfate, 5 µg of pC111 DNA (Boehringer Mannheim, Mannheim, Germany) and 3 µg of sonicated salmon sperm DNA, in a volume of 10 µl. Post-hybridization washing was at 42°C in 2x SSC–50% formamide (×3) followed by three washes in 0.1x SSC at 60°C. Biotin-labelled DNA was detected with Cy3-conjugated avidin.
REFERENCES


