Recent human-specific spreading of a subtelomeric domain.

S Monfouilloux, H Avet-Loiseau, V Amarger, I Balazs, C Pourcel, Gilles Vergnaud

To cite this version:

HAL Id: hal-01160639
https://hal-ensta-paris.archives-ouvertes.fr/hal-01160639
Submitted on 1 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Recent Human-Specific Spreading of a Subtelomeric Domain

Sylvaine Monfouilloux,* Hervé Avet-Loiseau,† Valérie Amarger,* Ivan Balazs,‡ Christine Pourcel,§ and Gilles Vergnaud*.† ¶

*Laboratoire de Recherche en Génétique des Espèces and † U.F. Cytogénétique Hématologique, Institut de Biologie des Hôpitaux de Nantes, 9, Quai Moncousu, 44035 Nantes Cedex, France; ‡ Lifecodes Corporation, 550 West Avenue, Stamford, Connecticut 06902; § Unité d’Immuno-Hématologie, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex, France; and ¶ Laboratoire de Génétique Moléculaire, Centre d’Études du Bouchet, B.P. 3, 91710 Vert le Petit, France

Received January 23, 1998; accepted May 1, 1998

INTRODUCTION

Humans and great apes are very highly similar at the DNA level despite the dramatic differences observed between these species. However, karyotype differences are visible, most of them associated with the ends of the chromosomes (Royle et al., 1994). A number of studies point to chromosome ends as hotspots for genome evolution, demonstrating differences even within the human population (Macina et al., 1995; Trask et al., 1998). Consequently the possibility that subtelomeric chromosomal rearrangements might have altered the regulation of one or more of these genes, which then had a significant impact on the evolution of our species, deserves investigation. For this purpose, and to develop efficient tools to study our recent evolution, human subtelomeric domains should be characterized in detail. Immediately adjacent to the telomere, in humans and in all eucaryotes investigated, begins a non-chromosome-specific subtelomeric domain. The size of the area located between the (TTAGGG)ₙ telomere repeats array and the chromosome-specific domain varies among chromosomes. It is only a few hundred basepairs long at human chromosome XpYpter (Baird and Royle, 1997), whereas it can be more than a hundred kilobases long at 16pter (Willkie et al., 1991) or 4qter (van Deukem et al., 1996). The transition from the subtelomeric, low-copy repeat domain to the chromosome-specific domain can be precisely defined, and in the few instances explored (16pter, 22qter, 4qter), genes are found at a high density starting a few kilobases proximal to the transition (van Deukem et al., 1996; Flint et al., 1997b; Wong et al., 1997). However, the succession of subtelomeric layers usually makes in situ investigations very difficult to analyze and interpret. Here we report the detailed description of a very recent instance of a highly successful chromosome end spreading in human, which interestingly may be sufficiently simple to provide a new tool to investigate our recent evolution.

MATERIALS AND METHODS

Minisatellite probes. Minisatellite pAC265 was previously reported (Ip et al., 1989). Minisatellite CEB92 was obtained by screening a commercial cosmid library by hybridization with the synthetic tandem repeat probe (Vergnaud, 1989; Vergnaud et al., 1991). 16C27 is (AGCTACGGTGTGGACT)ₙ.

Cell lines and DNA samples. Lymphoblastoid cell lines for individuals 1362-01, 1413-01, and 1413-02 and DNAs from the CEPH panel of 40 reference families were obtained from CEPH (Paris, France). Primate lymphoblastoid cell lines were purchased from ATCC (chimpanzee; Tank, Ref. 1847-CRL; orangutan; Puti, Ref. 1850-CRL).

Chromosome-specific cosmid libraries. High-density filters for human chromosome-specific cosmid libraries (chromosome 1, 6, and 17; library Nos. 112, 109, and 105) and individual cosmid clones were obtained from the Max-Plank Institute for Molecular Genetic (MPIMG). In addition, cosmid subclones generated from the half-telomeric YACs yRM2158 (chromosome 6qter) and B22 (chromosome 5qter) in the course of the development of chromosome-specific telomeric probes were a gift from Jonathan Flint (NIH and Institute of Molecular Medicine Collaboration, 1996).
Hybridization procedures. Hybridization was done as described in Vergnaud et al. (1991) using a hybridization oven.

Minisatellite isolation. Candidate minisatellite-containing fragments were identified by digesting cosmid DNA simultaneously with two of the frequent cutters AluI, HaeIII, and HinfI. After separation by agarose gel electrophoresis, fragments above 1 kb were extracted and hybridized to Southern blots of a few CEPH families. Fragments giving a polymorphic hybridization profile were subcloned into the PUC 18 vector for sequencing.

Segregation analyses in the CEPH families. The segregation of informative bands detected by minisatellite probes in the CEPH reference families was compared to the CEPH database version 8.1 (http://www.cephb.fr) by two-point lod-score analysis using CRIMAP version 2.4 (Green et al., 1990).

Cosmid walking and contig building. Successive directed walking was performed by hybridization using cosmid end probes on library filters. Cosmid end probes were generated by PCR elongation of primers flanking the cloning site of the vector (the T3 and T7 primers were used for YACs subcloned into the SuperCos1 vector; PL1 5′-ATACGACTCATATAGGGAG 3′ and PL2 5′-ACATACGATTAGGTGACAC 3′ were for cosmids based on the Lawrist4 vector (MPIMG-ICRF clones)). The reaction was done in 20 µl containing 1 µg of cosmid DNA; one primer at 0.5 µM; 25 µM dATP, dGTP, and dTTP; 0.75 µM dCTP (dNTPs from Pharmacia); 20 µCi of [32P]dCTP (ICN); and 2.5 u of Taq polymerase (Pharmacia) in 45 mM Tris, pH 9, 11 mM ammonium sulfate, 1.5 mM MgCl₂, 6.7 mM β-mercaptoethanol, 4.5 µM EDTA overlaid by oil. The reaction was performed with 30 cycles of 94°C for 20 s, 55°C for 30 s, and 72°C for 20 s in a PREM III thermocycler. Alternatively, when the resulting cosmid end probe was not appropriate for hybridization, usually because of the presence of a repetitive element, nearby restriction fragments were labeled by random priming.

DNA sequencing. Double-stranded DNA sequencing was done using the delta Taq sequencing kit (USB–Amersham) and end-labeled primers using kinase (New England Biolabs) and [γ-33P]ATP (Isotopchim). Sequence A1-4 was produced in parallel from different alleles by primer walking.

DNA sequence homology searches. Homology searches were performed using the BLAST server and DNA databases hosted by NCBI (http://www.ncbi.nlm.nih.gov).

Fluorescence in situ hybridization (FISH). Cosmids were labeled with biotin by nick-translation (Bionick kit; BRL) and hybridized to metaphase spread chromosomes of lymphocytes from the different cell lines. The probe was denatured and annealed to human Cot-1 DNA (final concentration of 2 µg/200 ng probe). Hybridization was performed at 37°C overnight. The hybridization signal was detected using an avidin/anti-avidin antibody detection system by a fluorescence microscope.

RESULTS

Linkage Analysis of the Two Highly Polymorphic Multicopy Minisatellites CEB92 (DNF92) and pAC365 (DNF24) Demonstrates Coassociation at a Number of Chromosome Ends

Minisatellites CEB92 and pAC365 were independently isolated in the course of searches for minisatellites within the human genome (Ip et al., 1989; Vergnaud et al., 1991), and both detect patterns with multiple bands on Southern blots. The eight largest CEPH families were used for segregation analyses to...
deduce the chromosomal distribution of the loci detected by the two minisatellites. The pattern obtained in CEPH (Centre d’Etudes du Polymorphisme Humain) family 1362 is presented in Fig. 1. Eight bands, numbered 1 to 8, are detected in the father, Id 1362 01, and 10 bands numbered 13 to 22 are visible in the maternal lane, Id 1362 02. The chromosomal assignment deduced from two-point linkage analysis against published data available in CEPH database version 8.1, is indicated at the left- and right-hand sides. In family 1362, all the bands can be assigned to a chromosome, except band 8, which is not informative, and bands 6 and 16, which are a superposition of two alleles segregating independently in the progeny. Altogether, 125 segregations can be scored in the eight families using CEB92, all of which are identical to the segregating markers originating from a limited number of loci. In particular the most distal loci on chromosome 1 (p), chromosome 5 (q), chromosome 6 (q), and chromosome 17 (q) account for 95 of the 125 bands. Loci 9pter, 17pter, 8pter, and 11pter segregations account for 20 of the bands. Loci 9pter and 15pter are each suggested by only one segregation (Table 1).

Using pAC365 (DNF24) and following the same procedure, 101 bands can be analyzed. Five (corresponding to three loci) do not correspond to any pattern in the database. Seventy-three of the remaining 96 segregations correspond to the most distal locations typed at 1p, 5q, 6q, and 15q. Twenty-two coincide with the most distal informative loci on 6p, 8p, and 11p. Distant 9q is suggested by 1 band. The data obtained strongly suggest that the two loci are colocalized and linked on six chromosome ends (1p, 5q, 6q, 6p, 8p, 11p), with additional weak evidence for a 9q localization. In contrast, CEB92 (and not pAC365) is detected at a moderate frequency on 7pter and at high frequency at 17qter. Chromosome 15qter is the second most frequent pAC365 site (16 bands) but is suspected only once with CEB92 (Table 1). It also suggests that some of these sites are “major” sites on which one or both loci are usually present (1pter, 5qter, 6qter, 15qter, 17qter) and that the others (6p, 7p, 8p, 9q, 11p) are more rarely occupied. One explanation for these findings is that a duplication containing both loci is found at a number of chromosome ends, with the exception of chromosome sites 7p, 15q, and 17q, where the duplication is shorter for some reason.

**Table 1**

<table>
<thead>
<tr>
<th>Family</th>
<th>Child. No. of bands (a)</th>
<th>1pter</th>
<th>5qter</th>
<th>6pter</th>
<th>6qter</th>
<th>7pter</th>
<th>8pter</th>
<th>9qter</th>
<th>11pter</th>
<th>15qter</th>
<th>17qter</th>
<th>? (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F102</td>
<td>14</td>
<td>8(9)p/9(9)m</td>
<td>1.2/M</td>
<td>1.2/3.4</td>
<td>/3.4</td>
<td>/M</td>
<td>1.2/3.</td>
<td>/M</td>
<td>1.2/3.4</td>
<td>0P/0M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1331</td>
<td>11</td>
<td>9(8)p/8(8)m</td>
<td>1.3/M</td>
<td>1.2/3.4</td>
<td>/3</td>
<td>1.2/3.</td>
<td>1/3.</td>
<td>1.2/3.</td>
<td>/3.4</td>
<td>/M</td>
<td>1.2/3.</td>
<td>0P/0M</td>
</tr>
<tr>
<td>F1332</td>
<td>10</td>
<td>4(6)p/11(7)m</td>
<td>1.3/M</td>
<td>1.2/3.4</td>
<td>/3</td>
<td>1.2/3.</td>
<td>1/3.</td>
<td>1.2/3.</td>
<td>/3.4</td>
<td>/M</td>
<td>1.2/3.</td>
<td>0P/0M</td>
</tr>
<tr>
<td>F1347</td>
<td>10</td>
<td>8(10)p/5(9)m</td>
<td>1.3/M</td>
<td>1.2/3.4</td>
<td>/3</td>
<td>1.2/3.</td>
<td>1/3.</td>
<td>1.2/3.</td>
<td>/3.4</td>
<td>/M</td>
<td>1.2/3.</td>
<td>0P/0M</td>
</tr>
<tr>
<td>F1362</td>
<td>11</td>
<td>6(8)p/10(9)m</td>
<td>1.3/M</td>
<td>1.2/3.4</td>
<td>/3</td>
<td>1.2/3.</td>
<td>1/3.</td>
<td>1.2/3.</td>
<td>/3.4</td>
<td>/M</td>
<td>1.2/3.</td>
<td>0P/0M</td>
</tr>
<tr>
<td>F1331</td>
<td>15</td>
<td>11(11)p/11(11)m</td>
<td>1.3/M</td>
<td>1.2/3.4</td>
<td>/3</td>
<td>1.2/3.</td>
<td>1/3.</td>
<td>1.2/3.</td>
<td>/3.4</td>
<td>/M</td>
<td>1.2/3.</td>
<td>0P/0M</td>
</tr>
<tr>
<td>F1416</td>
<td>9</td>
<td>7(10)p/7(10)m</td>
<td>1.3/M</td>
<td>1.2/3.4</td>
<td>/3</td>
<td>1.2/3.</td>
<td>1/3.</td>
<td>1.2/3.</td>
<td>/3.4</td>
<td>/M</td>
<td>1.2/3.</td>
<td>0P/0M</td>
</tr>
<tr>
<td>F884</td>
<td>12</td>
<td>10(11)p/7(8)m</td>
<td>1.3/M</td>
<td>1.2/3.4</td>
<td>/3</td>
<td>1.2/3.</td>
<td>1/3.</td>
<td>1.2/3.</td>
<td>/3.4</td>
<td>/M</td>
<td>1.2/3.</td>
<td>0P/0M</td>
</tr>
</tbody>
</table>

**Note.** For each of the eight families, and for both minisatellites, the chromosomal assignments deduced from segregation analyses are indicated. The number of children available in each family is indicated (column 2). Column 3 indicates the number of informative bands on the Southern and (a) the total number of bands (including noninformative bands). For each site, 1, 2, 3, 4, m and p indicate the grandparental origin of the assigned bands: 1, paternal grandfather; 2, paternal grandmother; 3, maternal grandfather; 4, maternal grandmother; p, paternal band of unknown origin; m, maternal band of unknown origin (unknown origin usually reflects the presence of a band with identical size in both grandparents; all families except F102 are three-generation families). The rightmost column reports the number and origin of unassigned bands. In some individuals, two bands are unassigned but correspond to a single locus (size in both grandparents; all families except F102 are three-generation families). The rightmost column reports the number and origin of uninformative bands. In some individuals, two bands are unassigned but correspond to a single locus (size in both grandparents; all families except F102 are three-generation families). The rightmost column reports the number and origin of uninformative bands. In some individuals, two bands are unassigned but correspond to a single locus (size in both grandparents; all families except F102 are three-generation families). The rightmost column reports the number and origin of uninformative bands.
telomeric half-YACs (chromosomes 5 and 6; NIH and Institute of Molecular Medicine Collaboration, 1996). The walk in one direction quickly reached an end as no clone could be obtained from any source after a single step. Restriction maps in the vicinity of CEB92 were identical when taking into account minisatellite allele size differences, suggesting that the different loci are highly homologous. Accordingly the EcoR1 restriction map could be used to easily identify and roughly locate breaks in homology of one contig compared to the others. The result of the walks is outlined in Fig. 2. The chromosome 17 restriction map becomes very different from the other three 10 kb away from CEB92 (defining Region 1). The chromosome 6 restriction map diverges from the chromosome 1–5 maps at 100 kb from CEB92 (Region 2). Finally, chromosomes 5 and 1 diverge 155 kb away from CEB92 (Region 3). Polymorphic monolocus minisatelites are located within 30 kb after the divergence point is reached on chromosomes 17, 6, and 5, and the segregation data obtained by screening the CEPH families confirm the chromosomal assignment to respectively 17qter, 6qter, and 5qter. The distance between CEB92 and the telomere is approximately 30 kb, as estimated by comparing the size of the 5qter YAC (B22; 225 kb; Kvaloy, 1993) to the distance between CEB92 and the cloning site in B22. This explains the failure to extend the walk in this direction and orients the four contigs with respect to the telomere. CEB92 and pAC365 are 60 kb apart and pAC365 is absent on chromosome 17, thus explaining the chromosome 17 linkage analysis results.

Structural Boundaries Defined by the Four Major Sites and Junction Sequencing

Junction 17q versus 1p, 5q, and 6q sites. The region distal to the junction and common to the four sites contains two minisatellite sequences separated by 540 bp and called CEB110 (most distal) and CEB102 (Fig. 2). The translocation occurred within the CEB102 minisatellite. The chromosome 17 allele contains an integer number of repeats, as deduced from the chromosome 17 linkage analysis results. The most proximal motif on the chromosome 5 allele is truncated and the adjacent one is partly deleted (Fig. 3A). A cosmid adjacent to the breakpoint (ICRFc105H11160, Fig. 2) detects a single locus on human chromosome 17qter by FISH analysis, confirming that the end of the subtelomeric domain has been reached on chromosome 17 (data not shown).

Junction 6q versus 1p and 5q sites. The breakpoint between the 6qter contig and the 1p–5qter contigs is localized 90 kb proximal in cosmids 6C5 and 6D1. The next cosmid, 6C6 (Fig. 2), detects a single signal on 6qter by FISH analysis. Three highly polymorphic monolocus minisatelites, CEB111, CEB113, and CEB114 (Cox et al., 1996), are present in cosmid 6C4, which also contains the proximal end of the YAC insert. Comparison between the cosmid restriction maps indicates the position of the translocation breakpoint. The telomeric sides of the 5qter and 6qter junction are perfectly identical along 80 bp until the breakpoint is reached (Fig. 3B), approximately in the middle of cosmid 6C5. To investigate whether the proximal half of cosmid 6C5 is chromosome 6 specific or still part of a subtelomeric region, the corresponding fragment as obtained by BssHII digestion of the cosmid (Fig. 2) was used for in situ investigations. Locus 6qter was detected in CEPH parents Id136201 and 141301 (Fig. 4), as expected. However, the probe also detected 8pter and 19pter. This demonstrates that the beginning of the chromosome 6-specific domain is not defined, in the present case, by the junction and that the probe defines an additional human subtelomeric layer, subsequently called region 4, which is common to 8p, 19p, and 6q and absent on 1p and 5q.

Junction 5q versus 1p. The boundary of the duplicated region on the 5qter locus is located 50 kb proximal to the chromosome 1–5 vs chromosome 6 divergence point. The junction sequences are aligned from telomere toward centromere in Fig. 3C, showing that the breakpoint is within the poly(A) tract (atypical poly(A) termination: C(A)7C(A)5C(A)5CACA) of an Alu repeat. The sequence CTTTACAG is present at both sides of the 5qter and 6qter junction. The telomeric half-YAC of chromosome 5q was interrupted between cosmids 6C5 and 6D4 by a gap of 60 kb. 1374 nucleotides are absent at the 17qter locus, some 6-specific domain is not defined, in the present case, by the junction and that the probe defines an additional human subtelomeric layer, subsequently called region 4, which is common to 8p, 19p, and 6q and absent on 1p and 5q.

**FIG. 2.** Physical map of the four major sites including the junction with the chromosome-specific domain. Restriction map of the translocated region in the loci 1pter, 5qter, 6qter, and 17qter. Solid black lines, translocated region; dotted lines, chromosome-specific domains. These maps are maps of single homologues. Distances from telomere, deduced from the chromosome 5 B22 half-YAC (Kvaloy, 1993), may be very different for other homologues due to chromosome end polymorphisms. Restriction map with rare-cutter enzymes Bs, BssHII; E, EagI; Rs, MluI; N, NotI; Nr, NruI; and Sf, SfiI and detailed map with current enzymes B, BamHI; Bg, BglII; E, EcoRI; H, HindIII; and S, Smal. A few HaellII (Ha) sites quoted in the text are also indicated. Cosmids used for in situ hybridization are indicated as solid lines. The 6qter contig is interrupted between cosmids 6C4 and 6D4 by a gap of 60 kb. 1374 nucleotides are absent at the 17qter locus and are indicated by Del1.4. Minisatelites are drawn as boxes: CEB92 (Accession No. (Ac) Y13539), CEB110 (Ac Y13547), CEB102 (Ac Y13546, Y13547, Y13548), CEB111 (Ac Y13538). Other sequenced fragments are indicated by double arrows: Seqs A1 (chromosome 1p; Ac Y13540), A2 (chromosome 17q; Ac Y13546), A3 (chromosome 6q; Ac Y13545), and A4 (chromosome 5q; Ac Y13544). 2 kb sequenced for evaluation of the divergence (see Table 1); Seqs B (Ac Y13548) and B’ (Ac Y13540), 1pter- and 17qter-derived sequences overlapping the junction with the 17qter domain reported in Fig. 3A; Seqs C (Ac Y13549) and C’ (Ac Y13541), 1p- and 6q-derived sequences overlapping the junction with 6qter domain 4 as reported in Fig. 3B; and Seqs D (Ac Y13550) and D’ (Ac Y13542), 1p- and 5q-derived sequences overlapping the junction with 5qter as reported in Fig. 3C.
following the junction contains the YAC vector cloning site and is specific to human chromosome 5 as shown by FISH. The chromosome 5qter contig is built from cosmids subcloned from a YAC and was not confirmed by an independent genomic library resource. Theoretically, the deviation from this chromosome's map could be the result of a cloning artifact in the half-YAC. This is unlikely because of the FISH data. More importantly, an artifact junction would not easily create the short duplication flanking the chromosome 5 Alu sequence and present only once at the chromosome 1 junction (Fig. 3C).

Sequence Similarity between the Four Major Sites

The EcoRI and rare-cutter restriction maps of the contigs are almost identical in the region of overlap, and the limited sequence analysis done at the distal side of the junction points further suggests that the different sites are highly similar.

To precisely measure the sequence similarity between the four different sites, the sequence corresponding to the 2 kb just distal to CEB110 (Fig. 2) was determined for one chromosome 1 and two chromosome 5, 6, and 17 independent alleles (Accession Nos. Y13542 to Y13547; more detailed information can also be found at our site, http://www.igmors.u-psud.fr/iech). The sequence on the chromosome 17 allele is only 639 bp long due to the absence of 1374 nucleotides. The divergence between the heterologous chromosome sites is approximately 1% (0.56–1.44% for 2034 bp and 0.2–1.8% for 639 bp, P < 0.05) and is not significantly higher than the divergence between the homologous alleles, suggesting a divergence time of not more than 1–2 million years for the major sites (Efstratiadis et al., 1980).

Analysis of Putative CpG Islands

CpG islands appear to be good indicators of the presence of genes in telomeric (GC rich) domains as shown by Flint et al. (1997b). CpG islands are often revealed by the aggregation of recognition sites for the rare-cutter enzymes NotI, Eagl, BssHII, KspI, Mlul, and Nrul. The mapping of the corresponding sites across the duplicated domains reveals two such aggregates. The first is localized 8 kb proximal to CEB92 (Fig. 2) and comprises Eagl, BssHII, and KspI sites. The second aggregate is the coincidence of a Nrul and BssHII site in Region 4 immediately proximal to the chromosome 6/chromosomes 1 and 5 junction.

To investigate the nature of the first potential CpG island, a 4-kb fragment was sequenced from a chromosome 5 allele. The proximal half of the 4-kb fragment contains the two minisatellite tandem repeats CEB110 and CEB102 (Seq A1, Fig. 2). The repeat unit in CEB110 is 69 nucleotides long and contains one Eagl site, thus creating a cluster of Eagl sites over a short distance. The CEB102 repeat unit is 105 bp long (Fig. 3A) and the sequenced fragment contains only the distal end of the tandem array. The 540 bp separating CEB110 and the more proximal CEB102 contain the Eagl, BssHII, and KspI rare-cutter sites. This segment is very GC rich (68% GC) but with a low (0.54) observed/expected CpG dinucleotide ratio. No significant homology of this fragment with CpG islands or EST sequences could be found.

A sequence database search identifies a strong homology with EST AA315353 in the distal half of the 4-kb fragment that was sequenced on different alleles to measure the site similarities. More precisely, sequence from 265 to 367 of the EST sequence is flanked in the genomic sequence by consensus splicing sites. The homology is within the 1374-bp region absent on chromosome 17. The 103-nucleotide homology is perfect with alleles from 1pter, 5qter, and 6qter. No other part of EST AA315353 is seen in the 4-kb sequence generated and no other significant hit was obtained in the databases.

The BssHII site in the second candidate CpG island is 160 bp proximal to the chromosome 6/chromosomes 1 and 5 junction, and the Nrul site is 60 bp proximal to the BssHII site (Figs. 2 and 3B). A perfect sequence identity with EST AA315353 starts 14 nucleotides after the breakpoint on 6q and runs for 100 nucleotides. Identity between the three sequences stops abruptly at a consensus splice site on the cosmid (Fig. 3B). The rare-cutter enzyme sites of BssHII and Nrul (Fig. 2) are localized in this intron.

In contrast with the situation observed within the duplication, aggregates of rare-cutter sites are quickly encountered outside of the domain on chromosomes 5 and 6 at least. A NotI site is located on chromosome 6 25 kb proximal to the chromosome 6/1 and 5 junction, close to Eagl and Mlul sites (Fig. 2). The sequence of the proximal side of the NotI site starts at the NotI site at position 190 in coding sequence S78085 (PDCD2 gene) and stops abruptly at position 312. The genomic sequence at this position is the good consensus donor splice site GAG/gtgag. PDCD2 was independently assigned to 6q27, with no pseudogene (Kawakami et al., 1995), showing that it is part of the chromosome 6-specific domain and probably the very first gene on chromosome 6q 150 kb from the telomere.

In Situ Investigation in CEPH Reference Individuals

Id 1362-01, 1413-01, and 1413-02

To investigate the situation at the other sites suggested by linkage analysis, and to identify the ancestral site, we have undertaken a number of FISH investigations in CEPH reference individuals and in one chimpanzee and one orangutan cell line.

The three junctions define three subtelomeric layers, Region 1 (telomeric to the 17q junction), Region 2 (between the 17q and the 6q5q–1p junction), and Region 3 (between the 6q and the 5q1p junction). In addition, although not as well defined, a fourth subtelomeric domain is found between the chromosome 1–5/6 junc-
tion and the chromosome 6-specific domain. The size of Region 1 is estimated at 40 kb, Region 2, 90 kb, and Region 3, 60 kb. Region 4 is less than 25 kb long and is represented by the proximal half of cosmid 6C5 (see left-hand side in Fig. 4).

As expected, cosmids from Region 1 detect multiple loci in the human cell lines. The results obtained for Id 1362-01 are summarized in Fig. 4A. Region 1 is detected on both homologues at 1pter, 17qter, 6qter, 5qter, 7pter (but with a weaker signal at this site), and 8pter and on one homologue at 11pter and 15qter. Polymorphism for Region 1-associated sites is demonstrated by the absence of signal at 11pter or 15qter and the presence of a signal on both homologues at 6pter, 19pter, and 19qter in Id 1413 01 (Fig. 4C). Except for 19pter and 19qter, all the sites detected by Region 1 cosmids were suggested by linkage analysis. On a per individual basis, however, linkage analysis expectedly misses some sites (such as 7pter, 8pter, and 15qter in Id 1362-01), presumably because some of the bands detected on the Southern blots are not informative (Fig. 1 and Table 1). As expected, cosmids from Region 2 do not detect 17qter. Cosmids from proximal region 2 containing minisatellite pAC365 (Fig. 2) fail to detect 7pter in all three CEPH individuals, in agreement with linkage data (Table 1), suggesting that 7pter contains approximately the distal half of Region 2. Similarly, the sudden loss of signal on 15q suggests that the chromosome 15q translocation junction is at most 20 kb distal from the 6pter/1p–5qter junction in Region 2.

Cosmids from Region 3 do not detect 8pter, suggesting that the breakpoint in homology on 8p is very close to the 6q junction. They stop detecting 6pter (observed in Id 1413-02, Fig. 4D) approximately 30 kb distal to the 5q–1p junction.

The probe defining Region 4 detects 6qter, 8pter, and 19pter in both Id 136201 and 141301, which confirms the similarity of the 6q and 8p alleles. Region 4 is present on chromosome 19p in Id 136201 despite the fact that Region 1 is absent at this site in this individual (Figs. 4B and 4C).

Region 1 detects both chromosome 9qter homologues in individual 141302, in agreement with linkage data, but with a weaker signal (a behavior similar to 7pter). However, proximal Region 2, which contains minisatellite pAC365, is not detected in this individual by in situ, in disagreement with linkage data. The reason for this discrepancy is unknown and could reflect additional rearrangements at this site. The in situ investigation in this restricted set of individuals demonstrates the existence of at least three different chromosome 9qter alleles.

In Situ Investigation in Chimpanzee and Orangutan

Cosmids from Regions 2 and 3 detect a number of loci, both telomeric and interstitial, in chimpanzee (Fig. 4D). The interstitial sites, 1q42, 4q27, 7p11, and 7q11 (human nomenclature), are also detected in humans, and some of the telomeric sites are common. Cosmids proximal to the junction on chromosomes 17 and 6 are monolocus and at the same location in the two primates. In contrast, cosmid SA4, adjacent to the chromosome 5–1 junction on human chromosome 5, and monolocus at 5qter in humans and orangutan, is present at seven chromosome ends (1pter, 3pter, 4qter, 5qter, 6qter, 8pter, 10pter, and qter) and at an interstitial site (9cen) in chimpanzee (Fig. 4D).

DISCUSSION

By combining the use of segregation analyses in large families, chromosome walking, and in situ hybridization studies, we have characterized different layers constituting one example of subtelomeric duplications that is of interest in understanding the evolution of chromosome ends in higher primates. A number of specific features made this analysis possible. First, the translocation domain contains two very highly polymorphic minisatellites with a wide allele size range perfectly suited for Southern blot analyses (Fig. 1). This finding has facilitated the identification of the different sites and provided information complementing and strengthening the FISH and physical mapping investigations. Second, although it occurred very recently, as demonstrated by the extremely close sequence similarity of the different sites analyzed, the domain quickly spread to a significant number of chromosome ends. It has been possible in this study to identify this spreading clearly because this family of duplication events is capped by Region 1. Region 1 is monolocus on chromosome 17qter in both chimpanzee and orangutan, and it is most likely that this site is the ancestral site. This is further suggested by the structure of the Region 1/Region 2 junction. The chromosome 17 CEB102 minisatellite allele contains a perfect number of repeats, which is a feature common to 70% of human minisatellites, and both flanking sequences, the telomeric and centromeric side, share with the repeat units a high density of TG motifs (Fig. 3A). On the contrary, alleles at the other sites contain an imperfect number of motifs, the last motif is truncated, and the previous one contains an internal deletion. The two chromosome 17 alleles analyzed lack a 1374-bp fragment that is present on chromosome 1, 5, and 6 alleles and contains one exon of EST AA315353, another part of which is located 90 kb downstream on the chromosome 6 locus close to the second translocation junction within Region 4. Such split matches appear to be common events in the vicinity of telomeres (Nicholls et al., 1987; Flint et al., 1997a), and a third part of EST AA315353 is also found 10 bp proximal to the chromosome 4qter degenerate TTAGGG at the distal/proximal subdomain boundary.
It has been suggested (Flint et al., 1997a) that the subtelomeric domain is separated into a distal and a proximal part by a short stretch of degenerate TTAGGG motifs and that distal and proximal subdomains are not transferred together in a translocation event. In agreement with this view, the chromosomal...
distribution of the distal subdomain from the 6qter half-YAC yRM2158 as deduced from Macina et al. (1995) is different from the chromosomal distribution of the proximal part as described in the present study. Similarly, the telomere-associated sequence TelBam11 (Brown et al., 1990) is detected at many 17qter and 5qter alleles but not at 1pter or 6qter sites. However, it should be stressed that even a limited number of unequal subtelomeric translocations occurring at a low frequency during our genome evolution can produce a very intricate pattern of low-copy repeat regions. Regions 2, 3, and 4 are present, together or separately, at many other chromosome ends, in both human and chimpanzee. Some of the sites detected are interstitial, presumably reflecting events (such as chromosome end fusion) that occurred in a common ancestor. Data generated in different laboratories cannot be compared and discussed unless common, widely accessible reference individuals are used, such as CEPH parents 1362-01 and 1413-01, who correspond to two of the largest and most studied families in the making of human genetic maps. It would be even more significant, obviously, regarding studies aimed at unraveling the recent evolution of our chromosome ends and comparing distal and proximal subtelomeric domains, to use a collection of reference cell lines from different well-defined human populations. The reconstructed history of the spreading of Region 1, by dedicated studies using such resources, will have to be compatible with a number of observations made here. For instance, one possible sequence of events could be 17qter→7qter getting distal Region 2→15qter getting proximal Region 2→19qter getting Region 4 and creating the split match→6qter; 5qter→1pter. The most likely mechanism for the spreading is the presence of a cryptic balanced translocation in one parent followed by an unbalanced transmission to some of the progeny. This proposed succession of events makes the easily testable prediction, in comparison with the results regarding 7pter, presented in Trask et al. (1998), that Region 1 distribution is strikingly different in Pygmy populations (and perhaps even monolocus at 17qter).

Indeed the evolutionary significance of these translocations remains unclear. The definition of the subte-
lomeric domain as the area between the telomere itself and the start of chromosome-specific sequence is relative. Region 1 is monolocus in the two primates studied and not in human, whereas cosmid 5A4 is monolocus in both human and orangutan but not in chimpanzee. The very rapid establishment of new proximal subtelomeric domains could suggest that the domain confers some selective advantage. Chromosome 6qter is of particular interest because PDCD2 (Kawakami et al., 1995), a highly conserved gene, is located just proximal to the translocation. Riethman and colleagues (Macina et al., 1995) identified by RARE cleavage one 6qter allele 130 kb larger than the others in a collection of 11 individuals and therefore demonstrated that 6qter is polymorphic in the human population. We predict that this rare allele is the chromosome 6qter ancestor, devoid of Region 4. This, in association with the search for phenotypes associated with homozygosity for the rare allele, may provide an efficient tool for the study of the functional significance of subtelomeric domains.

ACKNOWLEDGMENTS

We thank Professors Jean Dausset and Howard Cann for the provision of DNA samples from the CEPH reference panel and of cell lines from selected individuals. We are very grateful to William Brown and Harold Riethman for providing telomeric YACs for respectively 5qter and 6qter and to Jonathan Flint for providing cosmid subclones from these YACs. We thank Mourad Sahbatou at CEPH for a final two-point lod score analysis of segregation data against CEPH ver8.1 data. More detailed data can be found at http://igmors.u-psud.fr/iech. This work was supported by grants from the EEC (EUROGEM Project EC Contract GENE-CT93-0101), from the French Ministry of Research (ACC-SV grant), and from the GIP-GREG. S.M. has been supported by a stipend from the Ministère de la Recherche and from the Association Française contre les Myopathies.

REFERENCES


