HAL
open science

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:

S Monfouilloux, H Avet-Loiseau, V Amarger, I Balazs, C Pourcel, et al.. Recent human-specific spreading of a subtelomeric domain.. Genomics, 1998, 51 (2), pp.165-76. 10.1006/geno.1998.5358 . hal-01160639

HAL Id: hal-01160639

https://ensta-paris.hal.science/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.

GENOMICS **51,** 165–176 (1998)
ARTICLE NO. GE985358

Recent Human-Specific Spreading of a Subtelomeric Domain

Sylvaine Monfouilloux,* Hervé Avet-Loiseau,† Valérie Amarger,* Ivan Balazs,‡ Christine Pourcel,§ and Gilles Vergnaud*,¶,1

*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'Etudes du Bouchet, B.P. 3, 91710 Vert le Petit, France

Received January 23, 1998; accepted May 1, 1998

The recent spreading of a subtelomeric region at nine different human chromosome ends was characterized by a combination of segregation analyses, physical mapping, junction cloning, and FISH investigations. The events occurred very recently in human genome evolution as demonstrated by sequence analysis of different alleles and the single location of the ancestral site at chromosome 17qter in chimpanzee and orangutan. The domain successfully colonized most 1p, 5q, and 6q chromosome ends and is also present at a significant frequency of 6p, 7p, 8p, 11p, 15q, and 19p ends. On 6qter, the transposed domain is immediately distal to the highly conserved, single-copy gene *PDCD2*. © 1998 Academic Press

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 (Wilkie *et al.*, 1991) or 4qter (van Deutekom *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 Deutekom *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 clones were obtained from the Max-Planck 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).

Sequence data for this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. Y13538 to Y13550.

[1] To whom correspondence should be addressed at present address: Institut de Génétique et Microbiologie, Bat 400, Université Paris-Sud, 91405 Orsay Cedex, France. Telephone: +33 1 69 15 62 08. Fax: +33 1 69 15 66 78. E-mail: gilles@igmors.u-psud.fr.

0888-7543/98 $25.00
Copyright © 1998 by Academic Press
All rights of reproduction in any form reserved.

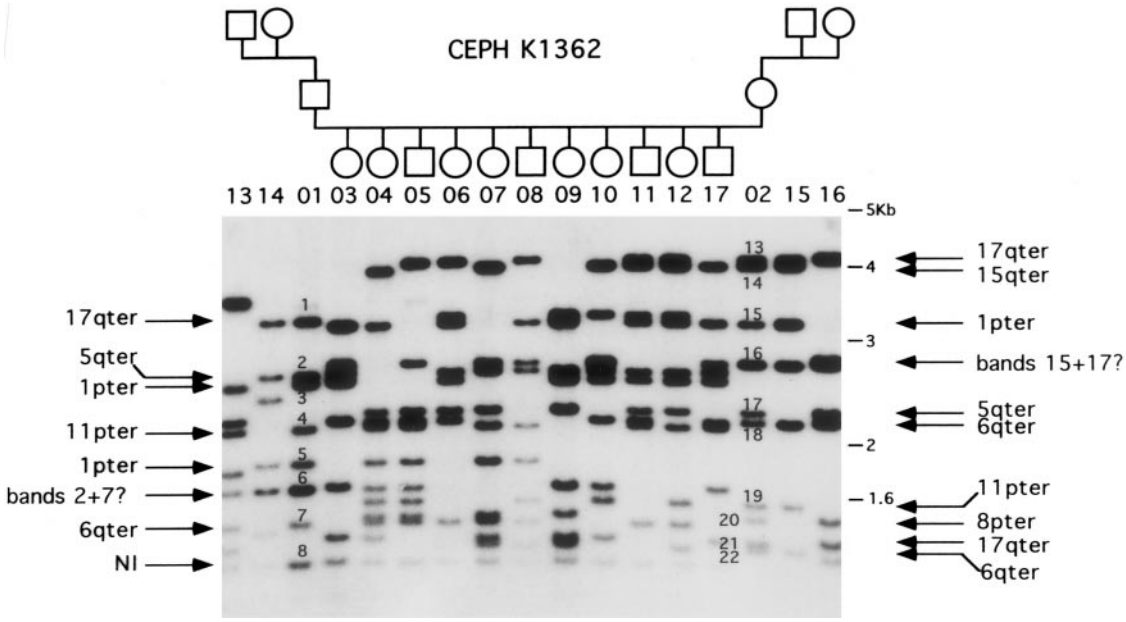


FIG. 1. Multilocus distribution of minisatellite locus CEB92. Southern blot analysis: Hybridization profile of CEB92 on CEPH family 1362 (*Hae*III digest). The size of each band is indicated in kilobases. CEB92 reveals 8 paternal (Id136201) and 10 maternal (Id136202) bands. Segregation of 6 of the paternal alleles was easily read, and a significant linkage with one terminal CEPH database marker was found for each allele. The CEPH version 8.1 segregation database is accessible at http://www.cephb.fr. The localization deduced from two-point lod-score analysis is indicated at the left. For convenience, in all families analyzed, bands of paternal origin are numbered in the range 1 to 12, in decreasing size order, whereas maternal bands are numbered starting at 13. Paternal band 6 has a double intensity compared to other bands of similar size, suggesting a superposition of two bands. The segregation of the band intensity (no band, normal intensity, double intensity) in the progeny is compatible with band 6 being the superposition of two bands allelic of band 2 and band 7. Band 8 has a double intensity in the father and is present at a normal intensity in all children, suggesting that it corresponds to one monomorphic locus (noninformative; NI). Nine segregating maternal alleles were linked to terminal markers with localization as indicated on the right. The maternal band 16, like paternal band 6, can be interpreted as the superposition of bands allelic to 15 and 17.

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 *Alu*I, *Hae*III, and *Hin*fI. 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′ ATACGACTCACTATAGGGAG 3′ and PL2 5′ ACATACGATTTAGGTGACAC 3′ were for cosmids based on the Lawrist4 vector (MPIMG-ICRF clones)). The reaction was done in 20 μl containing 1μg of cosmid DNA; one primer at 0.5μM; 25μM dATP, dGTP, and dTTP; 0.75μM dCTP (dNTPs from Pharmacia); 20 μCi of [^{32}P]dCTP (ICN); and 2.5 u of *Taq* polymerase (Pharmacia) in 45 mM Tris, pH 9, 11 mM ammonium sulfate, 1.5 mM $MgCl_2$, 6.7 mM ß-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-^{33}P]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

TABLE 1

Segregation Data Analyses in the Eight Large CEPH Families

Family	Child.	No. of bands (*a*)	1pter	5qter	6pter	6qter	7pter	8pter	9qter	11pter	15qter	17qter	?? (*b*)
		DNF92 (CEB92)											
F102	14	8(9)p/9(9)m	1.2/M	1.2/3.4		/3.4	/M	1.2/		/M		1.2/3.4	0P/0M
F1331	11	9(9)p/8(8)m	1./3.4	1./3.4	/3.	1.2/3.	1./	1.2/3.		/.4		1.2/	0P/0M
F1332	10	4(6)p/7(11)m	1./3.4	1.2/3.	/3.		/3.4					1./.4	0P/0M
F1347	10	8(10)p/5(9)m	1./.4	1.2/3.4		1./3.	.2/			1./		1.2/.4	0P/0M
F1362	11	6(8)p/9(10)m	1.2/3.	.2/.4		1./3.4		/.4		1./3.	/3.	.2/3.4	0P/0M
F1413	15	11(11)p/10(11)m	1.2/3.4	1.2/3.4	1./3.4	1.2/3.4		1.2/	/3.			1.2/.4	0P/0M
F1416	9	7(10)p/7(9)m	.2/.4	1.2/3.4		1.2/3.				/.4		1.2/3.4	0P/0M
F884	12	10(11)p/7(8)m	1.2/3.	1.2/3.	.2/	1.2/3.	.2/	/3.4				1.2/3.4	0P/0M
Total (*c*)		63(74)p/62(74)	12.11	14.13	2.4	10.10	3.3	6.4	0.1	2.4	0.1	14.11	0P/0M
Total (*d*)		125(148)	23	27	6	20	6	10	1	6	1	25	0
		DNF24 (pAC365)											
F102	14	5(8)p/3(5)m	/M	P/		/M		1.2/		/M	1.2/		0P/0M
F1331	11	6(8)p/8(10)m	P/3.4	1./3.		1.2/3.4				/3.	1.2/3.		0P/1M
F1332	10	6(7)p/9(10)m	/M	1.2/M	/3.	1./3.4.				1./3.4	1.2/3.4		0P/0M
F1347	10	4(6)p/5(7)m	/M	/3.	/3.	1.2/3.4							2(1)P/0M
F1362	11	6(9)p/3(6)m	1./	P/		1.2/3.4				1.2/	/.4		0P/0M
F1413	15	7(9)p/10(12)m	1.2/.4	1.2/3.4	1./3.4	1.2/3.4			/3.	/.4	/3.		0P/0M
F1416	9	4(7)p/9(10)m	P/	1./3.4		1.2/3.4				/M	/M		0P/2(1)M
F884	12	7(8)p/10(10)m	1.2/3.	/3.		1.2/3.		/.4(*e*)		1./M	1.2/3.4		0P/0M
Total (*c*)		45(62)p/56(70)m	7.7	8.8	1.4	13.14	0.0	2.4	0.1	4.7	8.8	0.0	2(1)P/3(2)M
Total (*d*)		101(132)(*e*)	14	16	5	27	0	6(*e*)	1	11	16	0	5(3)

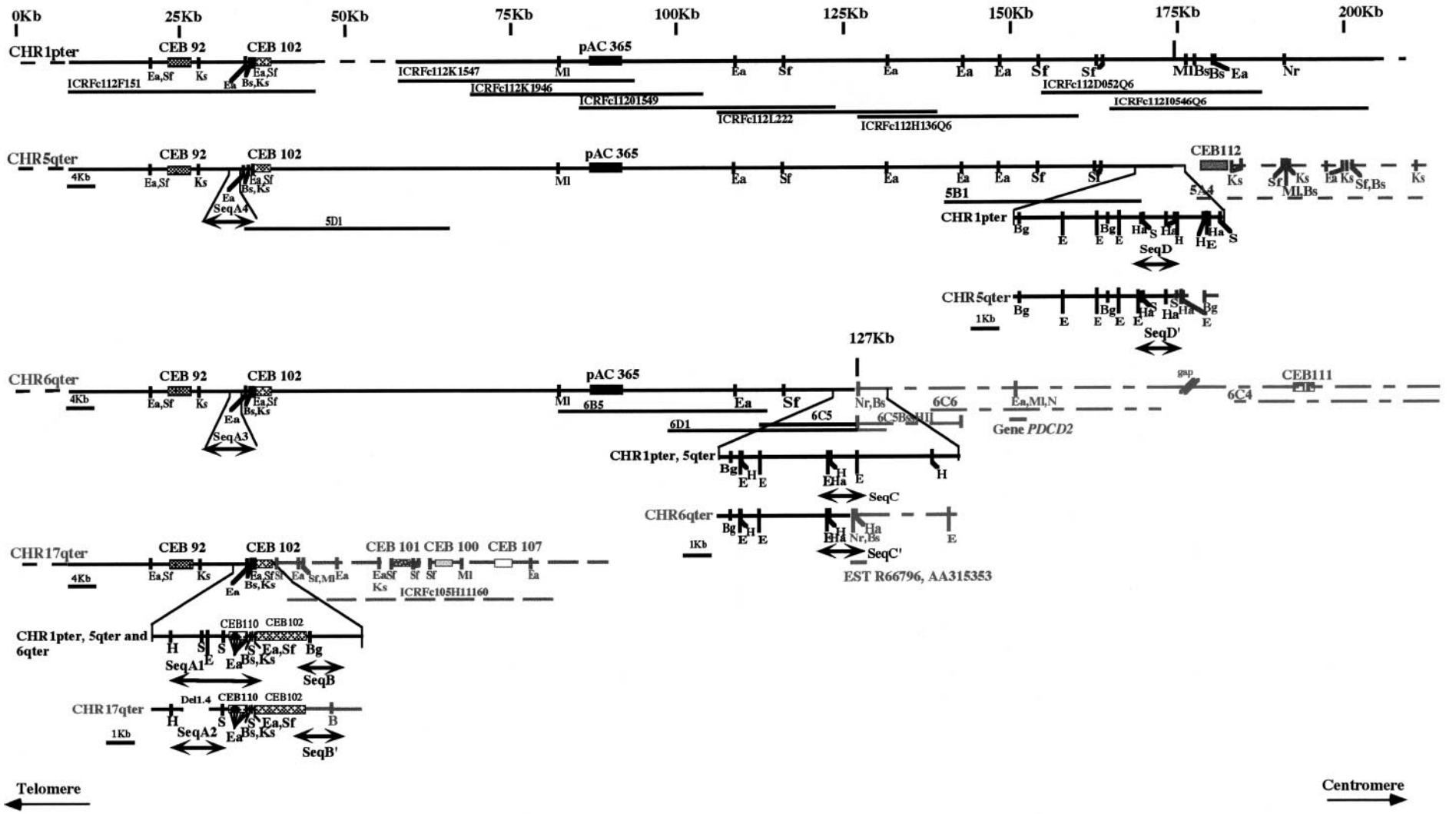
Note. For each of the eight families, and for both minisatellites, the chromosomal assignments deduced from segregation analyses are indicated. The number of children available in each family is indicated (column 2). Column 3 indicates the number of informative bands on the Southern and (*a*) the total 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 (*b*). Total (*c*) reports the total number of bands analyzed, taking into account parental origin. Total (*d*) cumulates the data. In one instance (DNF24, F884), four bands coming from the maternal grandmother segregate like 8pter loci (*e*), suggesting that in this individual proximal region 2 has been amplified fourfold.

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 segregation from at least one CEPH database marker. The cosegregating markers originate 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 6pter, 7pter, 8pter, and 11pter segregations account for 28 of the bands. Loci 9qter and 15qter 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. Distal 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 additionally 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.

Structural Analysis of the Four Major Sites, 1pter, 5qter, 6qter, and 17qter

Cosmid contigs corresponding to the four major CEB92 sites were built in parallel starting from CEB92 by screening chromosome-specific libraries (chromosomes 1, 6, and 17) or cosmid subclones from


0Kb
25Kb
50Kb
75Kb
100Kb
125Kb
150Kb
175Kb
200Kb
CHR1pter
CEB 92
CEB 102
pAC 365
ICRFc112F151
ICRFc112K1547
ICRFc112K1946
ICRFc11201549
ICRFc112L222
ICRFc112H136Q6
ICRFc112D052Q6
ICRFc112I0546Q6
CHR5qter
CEB112
SeqA4
5D1
5B1
5A4
CHR1pter
SeqD
CHR5qter
SeqD'
127Kb
CHR6qter
SeqA3
6B5
6D1
6C5
6C6
6C4
CEB111
Gene PDCD2
CHR1pter, 5qter
SeqC
CHR6qter
SeqC'
EST R66796, AA315353
CHR17qter
CEB 101
CEB 100
CEB 107
ICRFc105H11160
CHR1pter, 5qter and 6qter
CEB110
CEB102
SeqA1
SeqB
CHR17qter
Del1.4
SeqA2
SeqB'
Telomere
Centromere

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 *Eco*RI 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 155kb away from CEB92 (Region 3). Polymorphic monolocus minisatellites 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 comparison with the distal end of the minisatellite. 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 minisatellites, 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 *Bss*HII 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 *Alu* repeat on chromosome 5 (underlined boldface on Fig. 3C), whereas no such repetition is found on the 1pter allele, suggesting that the 5qter allele is an ancestor of the 1pter allele. Cosmid 5A4

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, *Bss*HII; Ea, *Eag*I; Ks, *Ksp*I; Ml, *Mlu*I; N, *Not*I; Nr, *Nru*I; and Sf, *Sfi*I and detailed map with current enzymes B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; and S, *Sma*I. A few *Hae*III (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 6C6 and 6C4 by a gap of 60 kb. 1374 nucleotides are absent at the 17qter locus and are indicated by Del1.4. Minisatellites 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 Y13543), 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 *Eco*RI 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 alleles 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 *Not*I, *Eag*I, *Bss*HII, *Ksp*I, *Mlu*I, and *Nru*I. 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 *Eag*I, *Bss*HII, and *Ksp*I sites. The second aggregate is the coincidence of a *Nru*I and *Bss*HII 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 *Eag*I site, thus creating a cluster of *Eag*I 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 *Eag*I, *Bss*HII, and *Ksp*I 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 *Bss*HII site in the second candidate CpG island is 160 bp proximal to the chromosome 6/chromosomes 1 and 5 junction, and the *Nru*I site is 60 bp proximal to the *Bss*HII 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 *Bss*HII and *Nru*I (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 *Not*I site is located on chromosome 6 25 kb proximal to the chromosome 6/1 and 5 junction, close to *Eag*I and *Mlu*I sites (Fig. 2). The sequence of the proximal side of the *Not*I site starts at the *Not*I 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 6q/5q–1p junction), and Region 3 (between the 6q and the 5q/1p 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 6qter/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

Region 1 cosmids are monolocus at chromosome 17qter in both chimpanzee and orangutan, and Region 2 is absent at chromosome 17qter in both primates, which is identical to the situation in human.

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 5A4, 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, 3qter, 4qter, 5qter, 8qter, 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.

A

Telomeric side

```
 1pter ccagggatgcacgcagagtaaggatgtgtgtgtctacgcatgtgggggt
       ---------------------------------------------
17qter ccagggatgcgcgcagagtaaggatgtgtgtgtctacgnatgtgggggt

 1pter TGGGTNTGACGGGGTGTGTTCTGTGTGAGAACGTGTGTGTAGTGTTCACATGTCCTCTGTGCGTGAGTCCCTGTGTGTGATGTTGTGTTCTCGGTGTGAGTTCA
       --------------------------
17qter TNGGNGTGACGGGGTGTGTTCTGTGT...

 1pter TGGGTGTGACGGGGTGTGCTGTGTGTGAGAACGTGTGTGTAGTGTCCACATGTCCTCTGTGCGTGAGTCCCTGTGTGTGATGTTGTGTTCTCGGTGTGAGTTCA
 1pter TGGGTGTGA.......

 1pter TGGGTGTGACGGGGTGTGCTGTGTGAGAGAACATGTGTGTAGTGTTCACATGTCCTCTGTGCGTGAGTCCCTGTGTGTGATGTTGTGTTCTCGGTGTGAGTTCA
 1pter TGGGTGTGACGGGGTGTGCTGTGTGAGAGAACATGTGTGTAGTGTTCACATGTCCTCTGTGCGTGAGTCCCTATGTGTGATGTTGTGTTCTCAGTGTGAGTTCA
 1pter TATGTGTGACGGGGTGTGCTGTGTGAGAGAACGTGTGTGTAGTGTTCATATGTTCTCAGTGTGTGAG<------------deletion----------->TTCA
 1pter TGTGTGTGACGGGGTGTGCTGTGTGA--GAACCCGTGTGCAGTG
                                   1pter  agatctccccaaaggtagttcaaagctgggccctttcatttcccggatctaaccc
                                                                                        17qter....GTGTGAGTTCA
17qter TGGGTGTGACGGGGTGTGCTGTGTGAGAGAACATGTGTGTAGTGTCCACATGNCCTCTGTGCGTGAGTCCCTGTGTGTGATGCCGTGTTCTCAGTGTGAGTTCA
17qter tgtatgtgtgtagtgagtttatgctcttgtgtgtcagcacatgggtatgatgtgggtatgttctctctgtgggtg
```

Centromeric side

B

Telo: side

```
                                                                  60
5qter        CACAATATTTGCCAAAGTCAAGTTATCCATTACACTATTAATTTGTCATTCTT
     ------------------------------------------------------------
6qter AAAAATTCACAATATTTGCCAAAGTCAAGTTATCCATTACACTATTAATTTGTCATTCTT
                         PJ62
                                                                 120
5qter TTGTTTATATAGTCAATATCTCTATCTCAATTGGATCTATCTCAACTGCTTCTAAACAAG
      ----------------------^-^^^^^-^^^-^^^^^^^-^^-^-^-^^^-^^^^^^
6qter TTGTTTATATAGTCAATATCTCTGTTGACCTGCCAGTGTCTCCGCCGGTTGAAAGCGCGT
                                            -----------------------
                 EST1 (R66796) 5'end GTCTCCGCCGGTTGAAAGCGCGT
                       EST2 (AA315353) 5'end CTCCGCCGGTTGAAAGCGCGT
                                                                 180
5qter CCACCATAGTCTCTCCCATTTCAACAATCTCTTCCAAGTACCACTTCATTTCTTCTTTTC
      ^^^^^^^^^^^^^-^^^^-^^^^^^^^-^^^^-^^^^^^^^^^^^^^^^^-^^-^^^^^^
6qter GTCTGCGTCGGGTTCTGTTGGAGTGCGTTCGGTGTGCCGTGGGTC-GCGCTGCTTCCACC
            PJ61
      --------------    ----------------------------^--------------
EST1  GTCTGCGTCGGGTT    TGGAGTGCGTTCGGTGTGCCGTGGGTCCGCGCTGCTTCCACC
EST2  GTTTGCNTCGGGTTCTGTTGGAGTGCGTTCGGTGTGCCGTGGGTCCGCGCTGCTTCCACC
                                                                 240
5qter ATATTTTTGAAAACTTTTGAAAAACTACCTATTTTCCTCCTCCATTTCTTGTTCATTCC
      ^^-^--^^^^^^-^^-^^^-^^^^-^^^^^^^-^^^^-^-^^^^^^^^^--^-^^-^--
6qter CAACTTCCTGTTAGGTAAGAGGCGCGTGAGGCTCCTGTGCCGGGGGCGGTGCTGCTCCC
      ---------------^^^^^-^--^-^^^--^--^^^^^^^^^^-^^-^^-^^^^^^^
EST1  CAACTTCCTGTTAGGACCCACCCGAGGACGGGTCACCATGGCTTTGGAGCACCTGGTAG
EST2  CAACTTCCTGTTAGCTGCAGGACAAGTTCGAGCATCTTAAAATGATTCAACAGGAGGAG
            BssHII                                                300
5qter ATTCTAGTGGACATGGAATCTGTTCCTCCTCCAAAACGGAATTTGGTCACCCTTAAATTA
      ^^^^^^-^-^^^^^^-^^^^^-^^^^---^^^^^^^-^-^^^^^^-^^-^^^-^^^^-^^
6qter GAGTCGGCGCGCGGCGGGGACGCGAGTCCGTAGGTGCTGGCGGGAGCGAGAGTCGGGTGG
      ^^^^^^^^-^^-^^^-^-^-^^-^--^^^^-^-^--^^^^^^^--^^--^^^-^--^^^^
EST1  TGTGGCATGTCCACAGTGAAGACCAGAGTTTCGTTGTCCTTAAGACTGACCTGGGGAGAC
EST2  ATAAGGAAGCTCGAGGAAGAGAAAAAACAACTGGAAGGAGAAATCATAGATTTTTATAAA
            NruI                                   HaeIII
5qter CTAAACCCAAAACAATATGTTGTCTTTATCTTTACCTCTCTGTGGCATTTAATGATAAGA
      ^^^-^--^^^^^^^^^^^^^^^^^-^^^^^^^^^^^^^^^-^^^^^-^^^^
6qter GGGACCCTCGCGAGCCCGCACTCCGCCTCTGGGTAGCAGCCTCTTCGGCC
      -^-^-^^^^-^^-^^^^^^^^^^^^^^^^^^^-^^---^^^^^^^^-^^^^^
EST1  GTGGCTGCAGGCATTTGAGGAAGACGGCACCGAAAGCGAAGGAAGCTCCTGCTCCTCCTA
EST2  ATGAAAGCTGCCTCTGAAGCACTGCAGACTCAGCTGAGCACCGATACAAAGAAAGACAAA
                                                 399
                                                EcoRI
5qpter CCACTACTTTCTTCTCTTTTACCCTTCTTTCTTGAATTC
EST1  AAGCCGAAGCCAAAGTGAAGGCTTTAAAGGCCAAGAAGGCAGTGTTTGAAAGGTGTTCCG
EST2  CATCCTCTGCTGCCCCCTGCCCCATACACGTGATGGAGCAGAAAACGTGCTGTGTGAACC
EST2  TGTGACTTCAGGGCCTGTTGACGTGGTCGTGCTTGCATACTCTCTGG
```

Centromeric side

```
C
Telomeric  side
      1                                                                  60
1pter GAAATAAGACTTTACAGCAGCCGGGTGCAGTGGTGCAGGCCTGTAATCCCAGCACTTTGG
      ------------------------------------------------------------
5qter GAAATAAGACTTTACAGCAGCCGGGTGCAGTGGTGCAGGCCTGTAATCCCAGCACTTTGG
                        ^------^--^----^^--^----------^-----------
                 HSALUCOL GGCCGGGCGCGGTGGCTCACGCCTGTAATCTCAGCACTTTGG
                                          HaeIII                 120
1pter CA-GCAGAGGCAGGCGGATCGCTTTGAGCTCAGGGCAACATAGCCAAAACCCCCCTCCcT
      --------------------^-----------------------------^----------
5qter CA-GCAGAGGCAGGCGGATCACTTTGAGCTCAGGGCAACATAGCCAAAAGCCCCCTCCNT
      ^-v--^-----^----------^^^---^-----^^^-^--vvv-^-^-^-^^---^^v-
(Alu) GAGGCCGAGGCGGGCGGATCAC---GAGGTCAGG---AGAT---CGAGACCATCCTGGCT
                                                                 180
1pter AGNNNACCCCCACCCCCTCCCTACCAAAAATACAAAA---CAGCAGGGCATGGTGGCGGG
      -vvvv-----------^-------------------------------------------
5qter ANNNNACCCCCACCCCGTCCCTACCAAAAATACAAAA---CAGCAGGGCATGGTGGCGGG
      -vvvv^^^^^^--------^----^------------vvv^---^----^----^-----
(Alu) AACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGCGTGGTAGCGGG
                                                                 240
1pter CGCCTGGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACCTGAACCCAGGAGG
      ------------------------------------------------------------
5qter CGCCTGGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACCTGAACCCAGGAGG
      -----v----------------^--------------------^^-^-------^-----
(Alu) CGCCT-GTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAGG
                                                                 300
1pter CAGACATTGCAGTGAGCCAAGATCACGCCACTG----CCAGCCTGGATGACAGAGCAAGA
      ------------------------------------------------------------
5qter CAGACATTGCAGTGAGCCAAGATCACGCCACTG----CCAGCCTGGATGACAGAGCAAGA
      -^--^^------------^-----^--------vvvv---------^^--------^---
(Alu) CGGAGCTTGCAGTGAGCCGAGATCGCGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGA
                                                 345
*hscb7        CAAAAAAACAAAAACAAAAA--CAAA--TCAGAATGGGA
1pter CTCCACCTCAAAAAAACAAAAACAAAAACACAAACCTCAGAGCAC
      ---------------------------------^^^^^^^^^-^--
5qter CTCCACCTCAAAAAAACAAAAACAAAAACACAGGTTAAGAGGGAC
      ----^^----------^-----^----
(Alu) CTCCGTCTCAAAAAAAAAAAAAAAAAA
                                                                 405
1pter CCCCAGGTATTCCAACCTAATCCTGGTGCCCCGCCTCTCACCACCCTTCTTCCTGTTTAA
      ----^^-^^--^^^-^^--^^^^-^^^^^^^^^^-^-^^-^^^^^^^^^-^--^^^^^^^
5qter CCCCGAGCCTTTACAGATACAAGTTTAAGAGGGACCTCTAAGCAAAAAATGCCAACCCTT

Centromeric  side
```

FIG. 3—*Continued*

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

FIG. 3. Sequence of the junction points between the four sites. (**A**) Sequence of the junction between 17qter and (1pter, 5qter, and 6qter). Seqs B (Ac Y13548) and B′ (Ac Y13540; see Fig. 2) are aligned from their telomeric ends. The common region is in black, the 17qter-specific area is gray. The telomeric flanking sequence of the minisatellite CEB102 is common to 1pter/17qter. Lowercase letters indicate sequence flanking the tandem repeat array. Repeat motifs are uppercased. Variant nucleotides between the repeats are gray. The *Bgl*II (bold, italicized) restriction site in the immediate chromosome 1pter flanking sequence has been localized in the detailed map of Fig. 2. The last unit at the 1pter minisatellite allele is shortest and is 44 nucleotides long instead of 105 nucleotides for a normal repeat. The minisatellite at the 17qter allele ends with a normal size repeat. The 17qter centromeric flanking sequence is very rich in TG motifs, reminiscent of the telomeric side and of the minisatellite repeat unit. (**B**) Sequence of the junction between the 6qter contig and the (1pter and 5qter) contigs. The sequences from 5qter (Seq C, see Fig. 2) and from 6qter (Seq C′, see Fig. 2) are aligned from the telomeric end. Sequences from the translocated domain (in black), 6qter-specific sequences (in gray), and EST1 (R66796) and EST2 (AA315353) (in gray). The consensus splice site at position 200 (bold, italicized) marks the end of the sequence identity between the ESTs and chromosome 6qter. Restriction sites indicated in the detailed map of Fig. 2 are italicized and underlined. Primers PJ61 and PJ62 have been used to sequence the junction point from one 1pter (ICRFc112L222) and one 6qter (ICRFc109C0311Q5) ICRF cosmid. The underlined sequence of EST AA315353 is found on Region 1 upstream of minisatellite CEB110 on chromosomes 1, 5, and 6 but is absent from chromosome 17. (**C**) Sequence of the junction between 1pter and 5qter loci. Sequences D and D′ (see Fig. 2) are aligned together with the most homologous *Alu* sequence found in the REPBASE collection of repeated elements: HSALUCOL. The *Alu* sequence is in gray. The direct repeats flanking the *Alu* element in the 5qter sequence are boldface and underlined. **Alu* element HSCB79B4 has a 3′ end homologous to the 1pter sequence, suggesting that the 1pter locus might have contained an *Alu* element before the translocation occurred (i.e., inter-*Alu*-element recombination event).

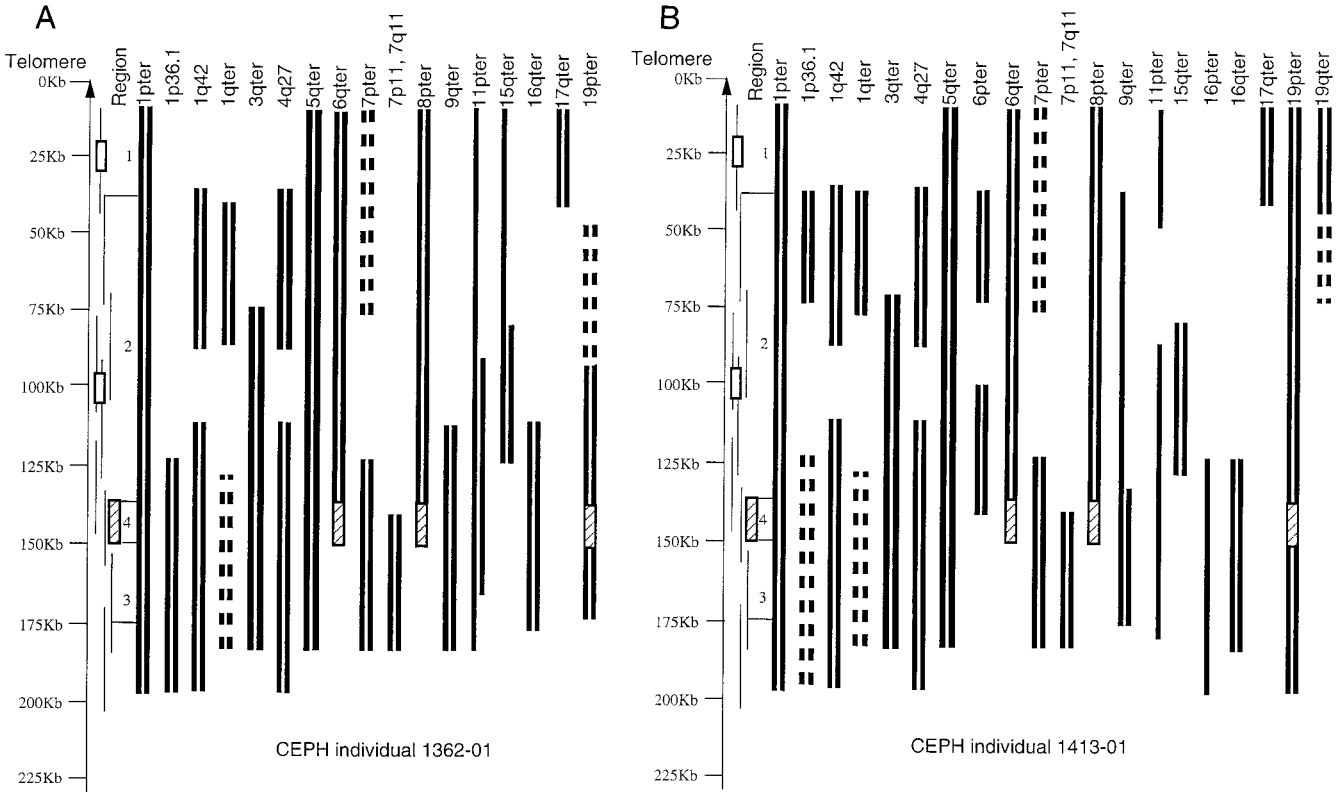


FIG. 4. Summary of the *in situ* investigations. Cell lines from CEPH parents 136201, 141301, and 141302 and from a chimpanzee were investigated *in situ* using cosmids ICRFc112F151 (Region 1), 5D1, ICRFc112K1547, ICRFc112K1946, ICRFc11201549 (Region 2), ICRFc112L222 (Regions 2 and 3), ICRFc112H136Q6, ICRFc112D052Q6, and ICRFc112I0546Q6 (Region 3) (Fig. 2). The hatched box represents the signal of subclone "6C5BsshII," shown in Fig. 2 and defining Region 4. Solid bars represent a supposedly continuous signal of overlapping cosmids for one chromosome. Interrupted lines (7pter, 9qter) represent weaker signals. Ten metaphases were observed for each cosmid. (A) Id 1362-01, (B) Id 1413-01, and (C) Id 1413-02.(D) Sites identified in chimpanzee (chromosome numbers refer to the numbers of the human homologues). The gray bars represent the signal obtained with cosmid 5A4, adjacent to the chromosome 5/1 junction, and monolocus at 5qter in both human and orangutan. Orangutan was investigated with Region 1 and Region 3 cosmids. Region 1 is monolocus at 17qter; Region 3 is observed at 3qter, 7p11, 7q11, and 7qter. The two cosmids adjacent to the junction at 17qter and 6qter are monolocus at the corresponding site in orangutan and chimpanzee.

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→7pter getting distal Region 2 →15qter getting proximal Region 2 →19pter 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-

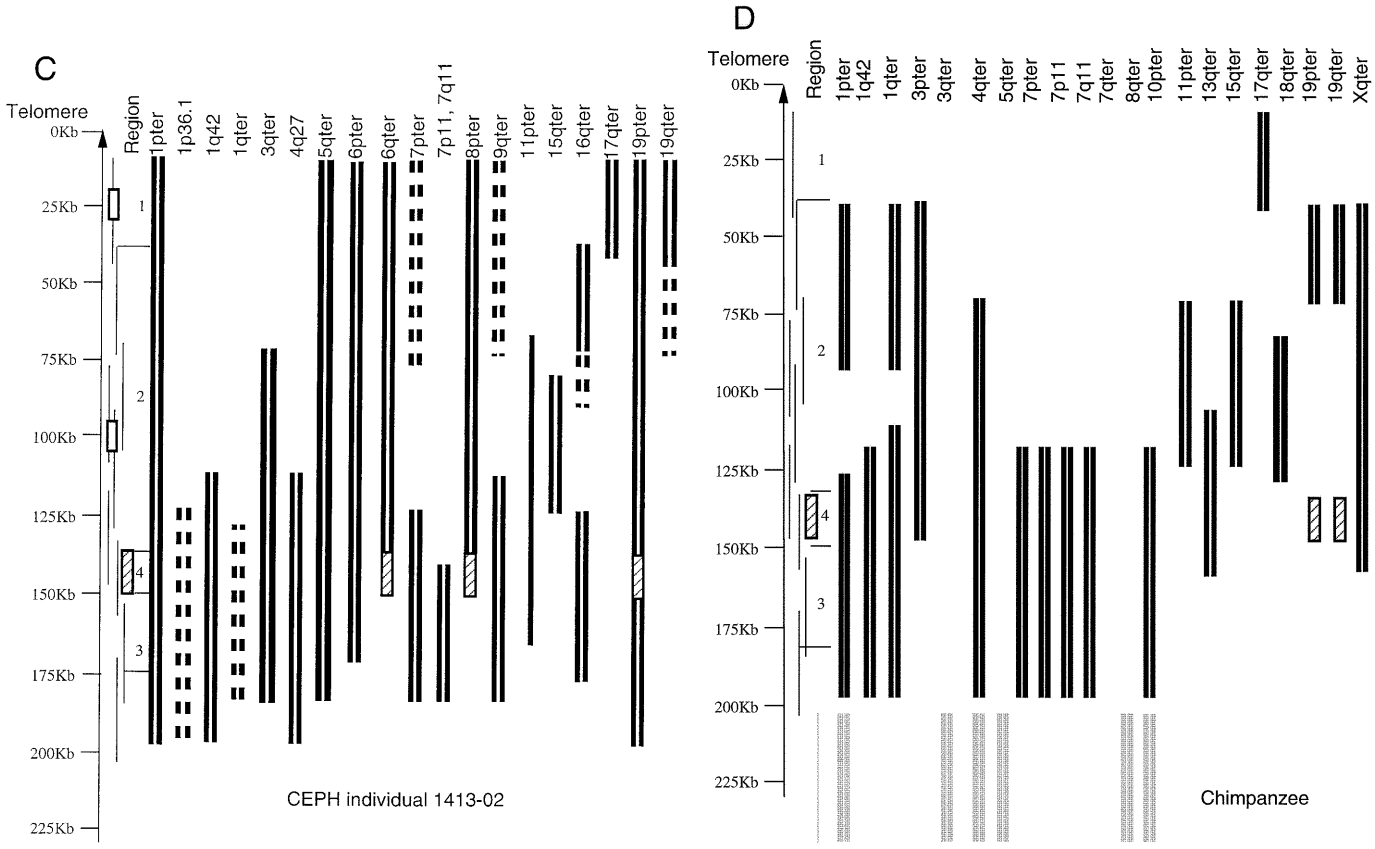


FIG. 4—*Continued*

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

Baird, D. M., and Royle, N. J. (1997). Sequences from higher primates orthologous to the human Xp/Yp telomere junction region reveal gross rearrangements and high levels of divergence. *Hum. Mol. Genet.* **6:** 2291–2299.

Brown, W. R. A., MacKinnon, P. J., Villasanté, A., Spurr, N., Buckle, V., and Dobson, M. J. (1990). Structure and polymorphism of human telomere-associated DNA. *Cell* **63:** 119–132.

Cox, S. A., Attwood, J., Bryant, S. P., Bains, R., Povey, S., Rebello, M., Kapsetaki, M., Moschonas, N. K., Grzeschik, K. H., Otto, M., Dixon, M., Sudworth, H. E., Kooy, R. F., Wright, A., Teague, P., Terrenato, L., Vergnaud, G., Monfouilloux, S., Weissenbach, J., Alibert, O., Dib, C., Fauré, S., Bakker, E., Pearson, N. M., Vossen, R. H. A. M., Gal, A., Mueller-Myhsok, B., Cann, H. M., and Spurr, N. K. (1996). European Gene Mapping Project (EUROGEM): Breakpoint panels for human chromosomes based on the CEPH reference families. *Ann. Hum. Genet.* 447–486.

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connel, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E.,

Shoulders, C. C., and Proudfoot, N. J. (1980). The structure and evolution of the human beta-globin gene family. *Cell* **21:** 653–668.

Flint, J., Bates, G. P., Clark, K., Dorman, A., Willinghan, D., Roe, B. A., Micklem, G., Higgs, D. R., and Louis, E. J. (1997a). Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Hum. Mol. Genet.* **6:** 1305–1314.

Flint, J., Thomas, K., Micklem, G., Raynham, H., Clark, K., Doggett, N. A., King, A., and Higgs, D. R. (1997b). The relationship between chromosome structure and function at a human telomeric region. *Nat. Genet.* **15:** 252–257.

Green, P., Falls, K., and Crooks, S. (1990). Documentation for CRI-MAP, version 2.4. Washington University School of Medicine, St. Louis, MO.

Ip, N. Y., Nicholas, L., Baum, H., and Balazs, I. (1989). Discovery of a novel multilocus DNA polymorphism [DNF24]. *Nucleic Acids Res.* **17:** 4427.

Kawakami, T., Furukawa, Y., Sudo, K., Saito, H., Takami, S., Takahashi, E., and Nakamura, Y. (1995). Isolation and mapping of a human gene (PDCD2) that is highly homologous to Rp8, a rat gene associated with programmed cell death. *Cytogenet. Cell. Genet.* **71:** 41–43.

Kvaloy, K. (1993). Ph.D. thesis, Oxford University.

Macina, R. A., Morii, K., Hu, X. L., Negorev, D. G., Spais, C., Ruthig, L. A., and Riethman, H. C. (1995). Molecular cloning and RARE cleavage mapping of human 2p, 6q, 8q, 12q, and 18q telomeres. *Genome Res.* **5:** 225–232.

Nicholls, R. D., Fischel-Ghodsian, N., and Higgs, D. R. (1987). Recombination at the human α-globin gene cluster: Sequence features and topological constraints. *Cell* **49:** 369–378.

NIH and Institute of Molecular Medicine Collaboration (1996). A complete set of human telomeric probes and their clinical application. *Nat. Genet.* **14:** 86–89.

Royle, N. J., Baird, D. M., and Jeffreys, A. J. (1994). A subterminal satellite located adjacent to telomeres in chimpanzees is absent from the human genome. *Nat. Genet.* **6:** 52–56.

Trask, B. J., Friedman, C., Martin-Gallardo, A., Rowen, L., Akinbami, C., Blankenship, J., Collins, C., Giorgi, D., Iadonato, S., Johnson, F., Kuo, W. L., Massa, H., Morrish, T., Naylor, S., Nguyen, O. T., Rouquier, S., Smith, T., Wong, D. J., Youngblom, J., and van den Engh, G. (1998). Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum. Mol. Genet.* **7:** 13–26.

van Deutekom, J. C. T., Lemmers, R. J., Grewal, P. K., van Geel, M., Romberg, S., Dauwerse, H. G., Wright, T. J., Padberg, G. W., Hofker, M. H., Hewitt, J. E., and Frants, R. R. (1996). Identification of the first gene (FRG1) from the FSHD region on human chromosome 4q35. *Hum. Mol. Genet.* **5:** 581–590.

Vergnaud, G. (1989). Polymers of random short oligonucleotides detect polymorphic loci in the human genome. *Nucleic Acids Res.* **17:** 7623–7630.

Vergnaud, G., Mariat, D., Apiou, F., Aurias, A., Lathrop, M., and Lauthier, V. (1991). The use of synthetic tandem repeats to isolate new VNTR loci: Cloning of a human hypermutable sequence. *Genomics* **11:** 135–144.

Wilkie, A. O. M., Higgs, D. R., Rack, K. A., Buckle, V. J., Spurr, N. K., Fischel-Ghodsian, N., Ceccherini, I., Brown, W. R. A., and Harris, P. C. (1991). Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome 16. *Cell* **64:** 595–606.

Wong, A. C. C., Ning, Y., Flint, J., Clark, K., Dumanski, J. P., Ledbetter, D. H., and McDermid, H. E. (1997). Molecular characterization of a 130-kb terminal microdeletion at 22q in a child with mild mental retardation. *Am. J. Hum. Genet.* **60:** 113–120.