Human Molecular Genetics, 1998, Vol. 7, No. 1 13–26

ARTICLE

Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes

Barbara J. Trask*, Cynthia Friedman, Antonia Martin-Gallardo[1], Lee Rowen, Carolyn Akinbami, John Blankenship, Colin Collins[2], Dominique Giorgi[3], Shawn Iadonato, Forrester Johnson, Wen-Lin Kuo[4], Hillary Massa, Tammy Morrish[5], Susan Naylor[5], Oanh T.H. Nguyen, Sylvie Rouquier[3], Todd Smith, David J. Wong, Janey Youngblom[6] and Ger van den Engh

Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195, USA, [1]Department of Biology, Universidad Autonoma de Madrid, 28034 Cantoblanco, Madrid, Spain, [2]Life Sciences Division, Lawrence Berkeley Laboratory, Berkeley, CA 94720, USA, [3]Centre de Recherches de Biochimie Macromoleculaire, CNRS, 34033 Montpellier, France, [4]Cancer Genetics Program, University of California at San Francisco Cancer Center, San Francisco, CA 94143, USA, [5]Department of Cellular and Structural Biology, University of Texas Health Sciences Center at San Antonio, TX 78284, USA and [6]Department of Biology, California State University–Stanislaus, Turlock, CA 95382, USA

Received August 11, 1997; Revised and Accepted October 21, 1997 DDBJ/EMBL/GenBank accession no. L78442

We have identified three new members of the olfactory receptor (OR) gene family within a large segment of DNA that is duplicated with high similarity near many human telomeres. This segment is present at 3q, 15q, and 19p in each of 45 unrelated humans sampled from various populations. Additional copies are present polymorphically at 11 other subtelomeric locations. The frequency with which the block is present at some locations varies among populations. While humans carry seven to 11 copies of the OR-containing block, it is located in chimpanzee and gorilla predominantly at a single site, which is not orthologous to any of the locations in the human genome. The observation that sequences flanking the OR-containing segment are duplicated on larger and different sets of chromosomes than the OR block itself demonstrates that the segment is part of a much larger, complex patchwork of subtelomeric duplications. The population analyses and structural results suggest the types of processes that have shaped these regions during evolution. From its sequence, one of the OR genes in this duplicated block appears to be potentially functional. Our findings raise the possibility that functional diversity in the OR family is generated in part through duplications and inter-chromosomal rearrangements of the DNA near human telomeres.

INTRODUCTION

The terminal bands of human chromosomes are unusual portions of the genome. These regions are particularly rich in genes (1). They show elevated meiotic recombination relative to the rest of the genome (2) and are the first regions to pair at the onset of meiosis (3). Special complexes form around the long arrays of T_2AG_3 repeats at the ends of all human chromosomes to guard against degradation and illegitimate fusions (4). The length of this telomeric repeat array varies somatically, and progressive shortening of the chromosomes correlates with reduced proliferative potential (5,6).

The transition from the common telomeric repeat to unique, chromosome-specific DNA is not abrupt. Instead, sequences that are repeated near the telomeres of many, but not all, chromosomes form a large transition zone, which is referred to as the subtelomeric region. With a few notable exceptions (7–15), such as the sequence of 284 kb at the telomere of 16p (14), the subtelomeric regions of human chromosomes are uncharted terrain. Although a handful of sequences, each cross-hybridizing

*To whom correspondence should be addressed. Tel: +1 206 685 7347; Fax: +1 206 685 7354; Email: trask@biotech.washington.edu

near the ends of a different subset of chromosomes, has been characterized (7,8,14–29), the overall size, structure, and gene content of these zones is largely unknown.

The subtelomeric zones exhibit an unusual form of polymorphism. Some individuals carry a given subtelomeric sequence on a particular chromosome, while others do not. All of the subtelomeric sequences that have been analyzed in more than one individual are polymorphic in copy number and location (17,18,23,25,27). The evolutionary pressures and processes that have shaped the subtelomeric regions and led to such variable and inter-related structures on different human chromosomes are not yet understood.

We have characterized a subtelomeric DNA duplication that provides some insight into the variability, complexity, and evolutionary history of these unusual regions of the human genome. Using a DNA segment cloned from chromosome 19, we demonstrate here that the blocks of DNA sequence shared by different chromosomes can be very large and highly similar. Our analyses of 45 individuals from different populations and three non-human primates show that these large-scale duplications are highly polymorphic in number and location and provide clues to their evolutionary history. While the frequency of duplication-containing alleles varies at some locations among human populations, three chromosomes appear to have contained the sequence before humans migrated around the world. In contrast to its multicopy distribution in humans, this subtelomeric block maps predominantly to a single locus in other primates. Of greatest significance is the discovery of three members of the olfactory receptor multigene family, one of which is potentially functional, within these large, polymorphically distributed duplications.

RESULTS

A cosmid derived from chromosome 19 cross-hybridizes to sequences near several telomeres

Our analyses of subtelomeres have centered around a 36-kb portion of chromosome 19 cloned in cosmid f7501. This clone mapped by fluorescence *in situ* hybridization (FISH) to both maternal and paternal homologs at 3qter, 15qter, and 19pter in the one individual studied previously (30). This individual also appeared to be a heterozygote for the presence of f7501-sequences at 11qter, because a FISH signal was observed at this site on only one homolog per cell.

Variation among human populations in the copy-number and locations of the 7501-sequence

To investigate normal variation in its distribution, f7501 was mapped by FISH to the chromosomes of 45 unrelated individuals sampled from eight ethnic groups. Metaphases of the three individuals shown in Figure 1A, B, and C illustrate some of the variation observed, and Figure 2 summarizes the results of all 45 individuals. Three locations, the termini of 3q, 15q, and 19p, cross-hybridized to this sequence in all individuals. All 90 homologs of chromosome 3q were positive for f7501. Virtually all copies of 15q and 19p also carry the f7501 block (87 and 89 of 90, respectively).

In addition to these very common sites, the f7501-block appears polymorphically near the tips of many other chromosomes. As a result of this variation, some individuals have only seven copies per cell; others have as many as 11. Figure 1A–C provides examples to illustrate the polymorphic distribution of this sequence. The cosmid hybridized to both homologs of 7p and 9q and one homolog at 8p in GM10471, a cell line established from a Biakan Pygmy (Fig. 1A). Another individual from the same population (GM10473) also carries the f7501 sequence on both homologs of 7p, but not on 8p or 9q (Fig. 1B). Instead, this individual has three copies on chromosome 16, near the q-terminus on both homologs and the p-terminus on one homolog. Another individual (978SK, Caucasian #8) showed hybridization to only one homolog each of 6p, 9q, 11p, and 16q in addition to the three common sites (Fig. 1C).

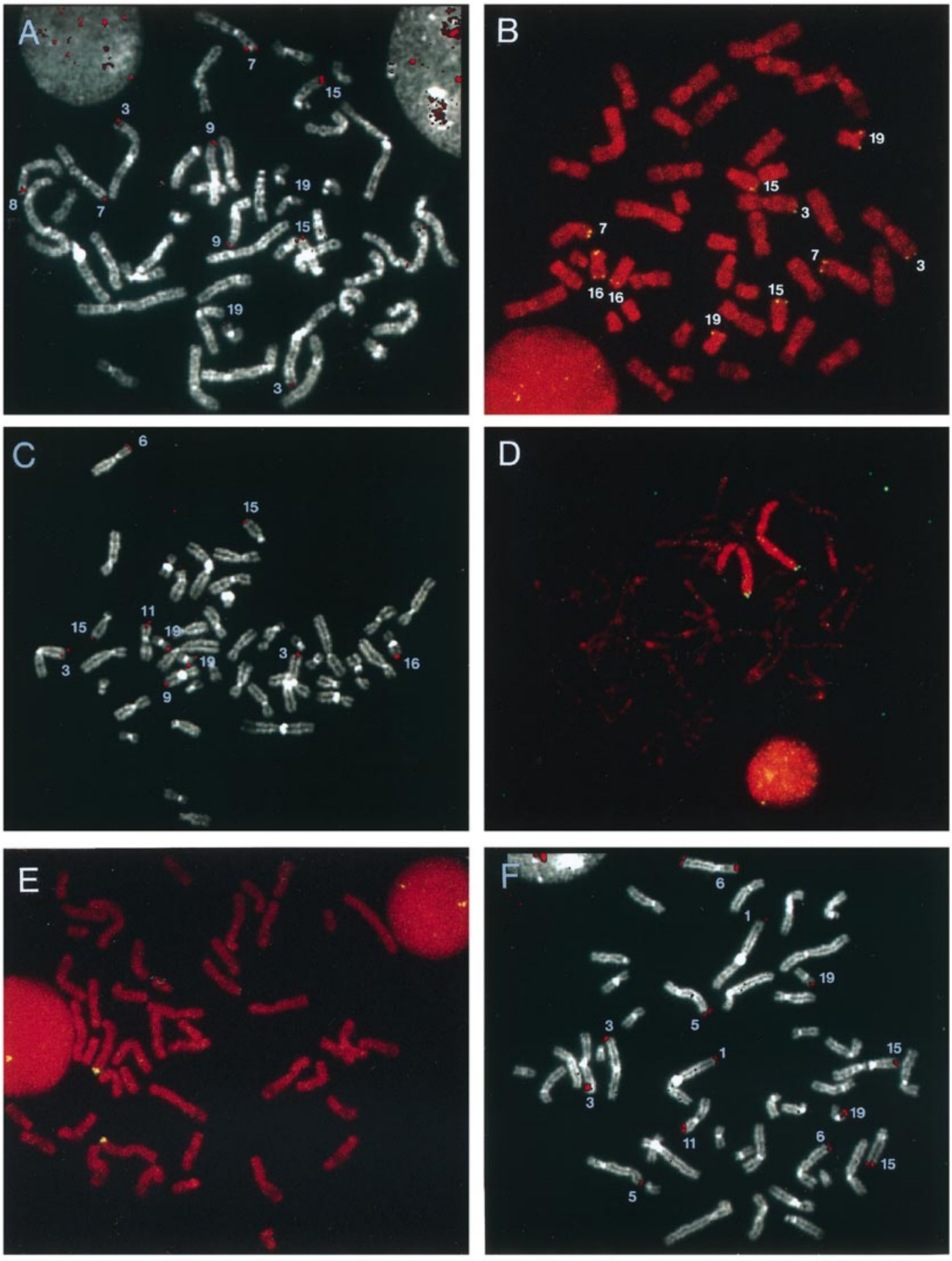
In the survey of 45 individuals, strong hybridization signals were observed at a total of 14 subtelomeric sites (large circles, Fig. 2), representing 29% of the 48 possible termini. Dim signals were infrequently noted on four additional sites (small dots, Fig. 2). With the exception of a single instance of dim interstitial labeling at 2q13–14, all sites were terminal.

We observed significant differences in f7501's chromosomal distribution among ethnic groups. The five representatives of the Biakan Pygmy population carry significantly more copies of the sequence than did several other groups (an average of 11 vs. 7.6 copies for the Kamerindians). The populations also varied in the frequency of 7501-positive alleles at some sites. The f7501-positive allele of chromosome 7p is more frequently present in the Biaka Pygmy sample than in the other populations. Only 13 f7501-positive 7pter alleles were detected in our study, and the overwhelming majority (11) belong to representatives of the African Pygmy tribes. The African Pygmies we sampled also carry the f7501-sequence more frequently on chromosome 16 (p and q) and less frequently on chromosome 11p than the other populations (*t*-test significance at 0.05 level).

Differences in the chromosome location of 7501-sequences among primates

In contrast to its multicopy distribution in humans, f7501-block hybridizes intensely to a single location in other primates (Fig. 2). It maps in chimpanzee and gorilla to 3qter (Fig. 1d, e), yet none of the 45 tested humans carries the f7501 sequence at 4qter, the orthologous location (31,32). Human chromosome 4, chimpanzee

Figure 1. (A–E) FISH patterns of cosmid f7501 on mitotic chromosomes of three unrelated humans (A–C), a chimpanzee (D) and a gorilla (E). The chromosomal distribution of P1 RMC0MP012 on cells from a fourth individual is shown in (F). The metaphase spreads were obtained from (**A**) GM10471, a cell line established from a Biakan Pygmy; (**B**) GM10473, from another Biakan Pygmy; (**C**) 978SK, a primary fibroblast culture of a Caucasian; (**D**) a PBL culture of *Pan troglodytes* (chimpanzee #2 in Fig. 2); (**E**) CRL1854 /ROK, a lymphoblast cell line of *Gorilla gorilla* (gorilla #1 in Fig. 2); and (**F**) GM1525, a cell line established from a Druze. The chromosomes consistently carrying hybridization signals are identified in (A–C) and (F). The f7501-positive site in chimpanzee and gorilla is at 3qter, which is orthologous to human 4qter. (A), (C) and (F) are pseudocolored digital images collected with a CCD camera; (B), (D) and (E) were photographed on color slide film. The probe sites are displayed in red in (A), (C) and (F) and yellow-green in (B), (D) and (E). The chromosomes were banded with DAPI in (A), (C) and (F) (gray) and with propidium iodide in (B) and (E). In (D), sequences homologous to human chromosome 4 were 'painted' with Texas-Red. The results of analyzing 10 metaphases per individual are summarized in Figures 2 and 4 for f7501 and P1 RMC0MP012, respectively.



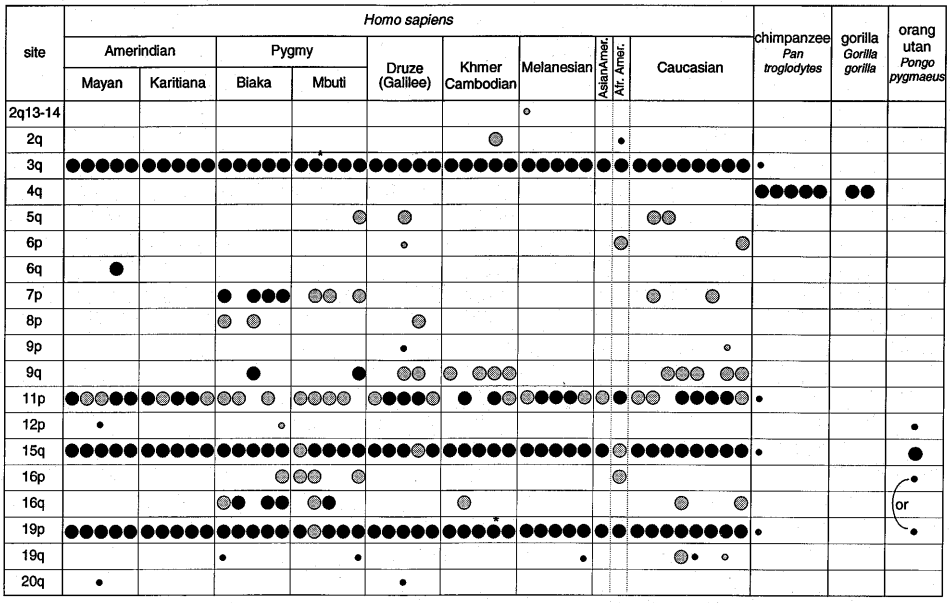


Figure 2. Chromosomal distribution of cosmid 7501 on 45 humans from various ethnic groups, five chimpanzees (PTR), two gorillas (GGO), and one orangutan (PPY). Each column summarizes the FISH data of 10 metaphases for one individual. Chromosomes are numbered according to the human (HSA) karyotype. The site on 2q13–14 labeled in a single chromosome is the only non-telomeric site detected. The symbols indicate that FISH signals were consistently observed on both homologs (black circles), on only one homolog per cell (gray circles), or not detected (no symbol). Small dots show the locations of dim sites seen infrequently on one (gray dots) or both (black dots) homologs. Segregation of hybridization with a specific homolog (gray) could be verified in several cases by taking advantage of size-polymorphisms at the centromeres (not shown). Probes that 'paint' sequences homologous to human chromosomes 4 and 15 were hybridized to chimpanzee and orangutan, respectively, to confirm the identity of the labeled chromosome in these species. The orangutan chromosome indicated as the ortholog of either HSA16 or 19 has not been identified by chromosome painting. The corresponding conventional nomenclature for the chromosomes labeled in non-human primates are PTR3, GGO3, and PPY16 (predominant sites) and PTR2, PTR9, PTR16, PTR20, PPY9, and PPY18 or20 (dim and infrequently detected sites). The identifiers of the human samples are, from left to right, GM10975-9, GM10965-9, GM10469-73, GM10492-6, GM11521-5, GM11373-7, GM10539-43, PHA-stimulated PBL cultures of eight anonymous donors, CGM1; and 978SK. The five chimpanzee samples are three PBL cultures, GM03542, and TANK. Gorillas 1 and 2 were a PBL culture and CRL1854/ROK, respectively. GM06213 was the source of orangutan chromosomes. Of the human samples, 22 were male and 23 were female. All non-human primates except one were male. The FISH results for GM10471, GM10473, 978SK, chimpanzee 2 and gorilla 1 are shown in Figure 1. Two chromosome rearrangements (indicated with *) involving the f7501 sequence were detected in these ostensibly normal cell lines. A translocation of the entire 3q-arm, including the f7501 signal, to 6p was noted in some GM10493 cells. In GM11376, the FISH signal on one homolog of 19p was significantly further from telomere on the other.

chromosome 3, and gorilla chromosome 3 are otherwise homologous using whole-chromosome paints (Fig. 1D and ref 31) and, with the exception of a pericentromeric inversion, by banding (32). In orangutan, the sequence is located on chromosome 16qter, the ortholog of human chromosome 15qter (31,32) (not shown). The FISH signals on these ape chromosomes were as intense and as efficiently detected as the FISH signals on the various human chromosomes, suggesting that they contain the bulk of the f7501 sequence. In one experiment (Chimp #1 in Fig. 2), we occasionally observed very small signals, which presumably represent only a small portion of the 7501 sequence, on the chimpanzee orthologs of human chromosome 3q, 11p, 15q, and 19p. These small signals lay internal to the AT-rich satellites that have been added to the ends of some chromosomes in chimpanzee since their divergence from our common ancestor (32). Dim signals were also observed infrequently on the orangutan orthologs of human chromosomes 12pter and either 16p or 19p (not shown).

The duplications on different chromosomes are large and very similar

In order to demonstrate that duplication of bulk of the 36 kb f7501 sequence accounts for the similarity in FISH signal at the different sites, we assayed for the presence of various portions of the sequence by FISH and PCR. In the first approach, subclones distributed across f7501 were mapped by FISH to six individuals (three to six subclones per individual). These subclones hybridized to multiple locations, closely paralleling the polymorphic pattern of the entire cosmid (not shown).

In the second approach, PCR primer pairs were designed from the sequence of the cosmid, which is described in more detail below. A panel of 24 monochromosomal hybrids, each containing a different human chromosome in a rodent background, was typed by PCR for 17 sites distributed across the f7501-sequence (Fig. 3a). Additional monochromosomal hybrid lines for chromosomes 11, 15, and 19 were also analyzed. As expected, chromosomes 3, 15, and 19 were positive, with a few minor exceptions, for all 17 PCR assays. The more infrequent 7501-positive alleles of other chromosomes apparently are not represented in this hybrid panel. None of three different hybrids containing chromosome 11 carries f7501-sequences detectable by PCR (Fig. 3a) or FISH (data not shown), although hybridization to 11pter was observed on 64% of the 90 homologs analyzed by FISH (Fig. 2).

Southern blot analyses confirmed the large size of the duplication and also revealed similarity in the gross structure of the duplicates (data not shown). Six probes were tested, each recognizing a different *Eco*RI fragment of f7501 (see Materials

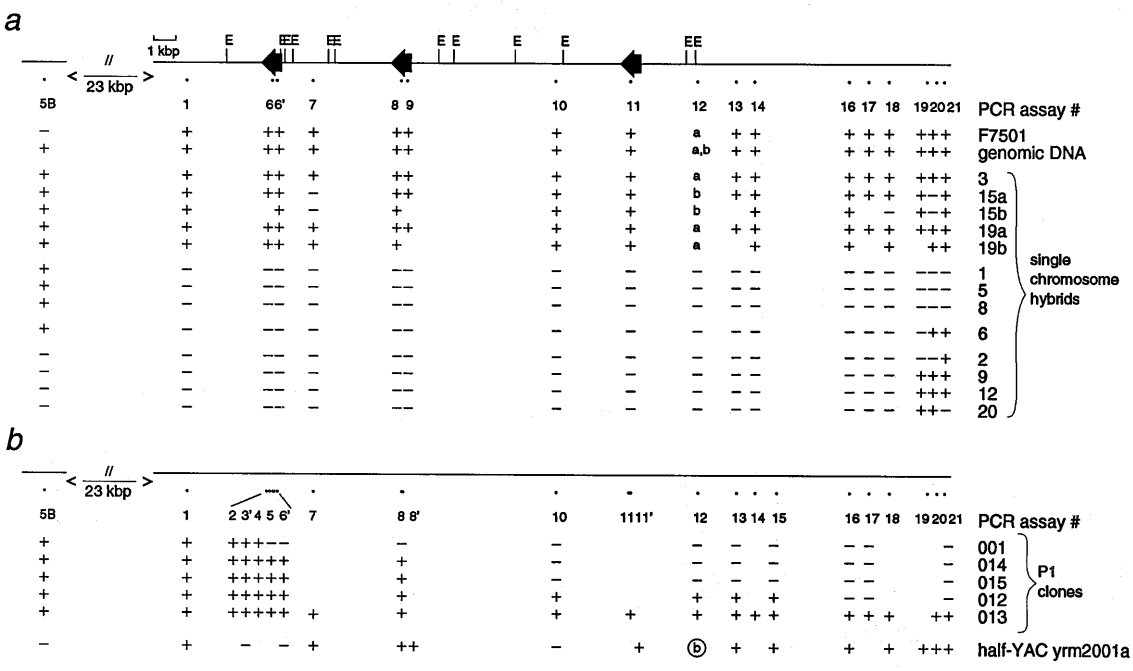


Figure 3. PCR typing of monochromosomal hybrids (**a**) and large-insert clones (**b**) with 17 primer pairs defined from f7501's sequence. The dots indicate the midpoint of the product amplified with each primer pair. The entire NIGMS panel #2 was tested for amplification with each primer pair, but only those chromosomes positive for at least one assay are indicated, +, positive; –, negative. (a) and (b) are two size-variants amplified using primer pair 12, which lies within the proximal end of half-YAC yrm2001a (27) (circle). The absence of a symbol indicates that the PCR test was not performed. The two chromosome 15 hybrids have in common their failure to amplify with two isolated primer pairs (7 and 20) and the same length polymorphism (form b) for assay 12. The large black arrows mark OR sequences in the cosmid (see text and Fig. 5). The RMC0MP prefix is omitted from the P1 names for simplicity.

and Methods). Four probes identified a single band in *Eco*RI-digested genomic DNA from individuals known to carry seven to 11 copies of the f7501 block. Only two probes detected restriction fragment length polymorphism among different chromosomes.

We compared the structure of paralogs on chromosomes 3 and 19 in more detail by characterizing two clones from a chromosome 3 cosmid library that overlap f7501's sequence. These clones, 176C1 and 214B2, overlap f7501 by ~34 kb and ~25 kb respectively. Remarkably, their *Eco*RI, *Bam*HI and *Hin*dIII restriction maps are identical to f7501's in their regions of overlap (not shown).

f7501 is flanked by sequences that are duplicated on additional chromosomes

In order to gauge the size of the duplication and the extent of polymorphic variation, we analyzed the chromosomal distribution of five P1 clones that overlap f7501. These clones were identified in a total genomic library by PCR screening with primer pair 6′. Therefore, the chromosomal origin of these ~70 kb segments is unknown. Overlap of the clones with f7501 was established by PCR (Fig. 3b) and by comparing the end-sequences of the P1s with f7501's sequence. One of the P1s, RMC0MP013, amplifies with all tested primer pairs and extends in both directions from f7501. Four P1s overlap f7501 by ~5.5–31.7 kb and have captured centromere-proximal sequence only (Fig. 4).

FISH analyses of these P1 clones on three individuals revealed that the f7501 block lies between sequences that are duplicated on more chromosomes than the f7501 block itself (Figs 1F and 4). Clones that include sequences proximal of f7501 (to its left in Fig. 4, indicated by the blue zone) cross-hybridize near the telomeres of chromosomes 1p, 5q, 6p, 6q, and 8p, in addition to the locations identified for f7501 in these individuals. These sites are highlighted in blue. P1 RMC0MP-013, which overlaps f7501 entirely and captures sequences both proximal and distal (green zone), hybridizes to all the locations seen with f7501 and the other P1s, but cross-hybridizes to six additional sites. These sites, the ancestral telomere–telomere fusion site at 2q13–14 (22) and near the termini of 6p, 9p, 12p, 20p, and 20q, which are highlighted in green, must share sequences that lie distal of f7501 on the chromosomes from which the P1 was cloned. Thus, a single walk from 7501 links the maps of 19 of the 48 human chromosome ends and an interstitial site at 2q13–14. Figure 4 further demonstrates that variants of chromosomes 9q and 11p differ by the presence or absence of the entire ~85 kb region spanned by these overlapping P1s.

This FISH pattern is consistent with the PCR results shown in Figure 3. Although most of the sequences across the f7501 block can be amplified from only chromosomes 3, 15, and 19 in the hybrid panel, flanking sequences are duplicated on additional chromosomes. The sets of positive chromosomes are different on either side. PCR primers in the most distal 2 kb of f7501's sequence amplify from chromosomes 2, 6, 9, 12, and 20, in addition to 3, 15, and 19. A primer pair defined from the end of a cosmid that overlaps f7501 and extends 23 kb off its proximal end (S. Iadonato, unpublished results) amplifies from chromosomes 1, 5, 6, and 8, in addition to 3, 15, and 19. It is notable that both the PCR and FISH results show that chromosome 6 usually

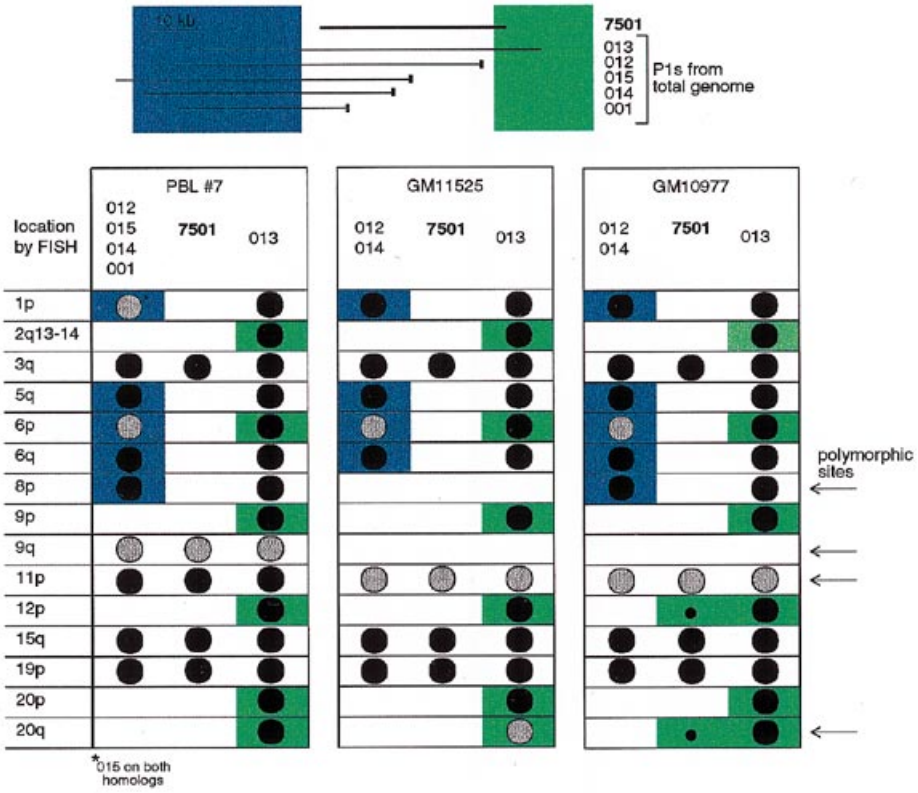


Figure 4. Summary of the FISH results of P1s that overlap 7501 to varying degrees on three individuals. F7501 maps to only a subset of the chromosomes identified by the flanking sequences.The lines at the top indicate the relative size and overlap of the P1s and f7501 (see Fig. 3b). The orientation of the contig is inferred by the overlap of f7501 and a half-YAC (Fig. 3b). See text for explanation of green and blue highlighting. The arrows mark polymorphic sites. The gray and black circles and dots are used as described in the legend to Figure 2. A metaphase spread of GM11525 hybridized with P1 RMC0MP-012 is shown in Figure 1F.

lacks the majority of the f7501 sequence, but contains sequences that map to either side of the f7501 block on other chromosomes.

Sequence analysis of f7501 reveals homologies to olfactory receptor genes

In order to learn the nature of the sequences that are tolerated with such variability in copy number and location, we determined the sequence of cosmid f7501 through a combination of random shotgun and directed sequencing strategies. Three regions with significant homology to olfactory receptor (OR) genes are the most notable features of the sequence (Fig. 5). Olfactory receptors are members of the larger family of G-protein coupled receptors, which have seven transmembrane segments (33). The coding regions of these genes are short, spanning only ~1000 bp, and intronless (33). These three regions lie in the same orientation and are denoted OR-7501A, OR-7501B, and OR-7501C. The three OR sequences in f7501 differ at the nucleotide level by 28–33% base-substitutions and 1–7% insertion/deletion changes.

The greatest similarity (49% identical and 70% similar amino-acids) was detected between OR-7501A and OR18, an OR gene expressed in rat (34) (Fig. 6). OR-7501A is most similar at the nucleotide level, to an OR gene expressed in dog (CfOlf4, U53682) (35) of the sequences currently in the public databases. It is also very similar to an OR gene expressed in dog testes (36) (DTMT, X64996) (42% identical and 66% similar amino-acids) and a taste-bud receptor expressed in rat (37) (TB641, U50949) (57% nucleotide identity, and 42% identical and 68% similar amino-acids).

OR-7501A appears potentially functional based on its sequence. It has an open reading frame of 306 amino acids and conserved sites that are known to be critical for function (33) (Fig. 6A). The regions that are the least conserved relative to published OR sequences correspond to variable transmembrane segments, which are thought to encode odorant-specificity (33). OR-7501B and OR-7501C are pseudogenes (Fig. 6B). OR-7501B has one frameshift caused by a single-base deletion and two internal stop codons. OR-7501C is disrupted by five in-frame stop codons and seven frameshifts (five caused by single-base insertions or deletions, one by an 8 bp deletion, and one by a 17 bp deletion). Excepting these defects, the proteins encoded by the three OR sequences would be 47–60% identical (Fig. 6B).

The cosmid's sequence can be oriented on the chromosome by virtue of its overlap with a sequence defined from the centromere-proximal end of a human half-YAC [yrm2001a (27)]. The half-YAC sequence is identical to bases 24691–24829 in f7501, except for a difference in the length of a variable $(T_{3-4}G)_n$ repeat within this overlap (n = 5 in the half-YAC and 6 in f7501). The length of this repeat also varies among chromosomes (Fig. 3). The half-YAC is positive for all PCR assays in the more distal portion of f7501's sequence. This fact orients the cosmid on the chromosome and places it ~50 kb [the size of the half-YAC (27)] from the end of at least one chromosome. This interpretation may be an oversimplification, however, because the half-YAC

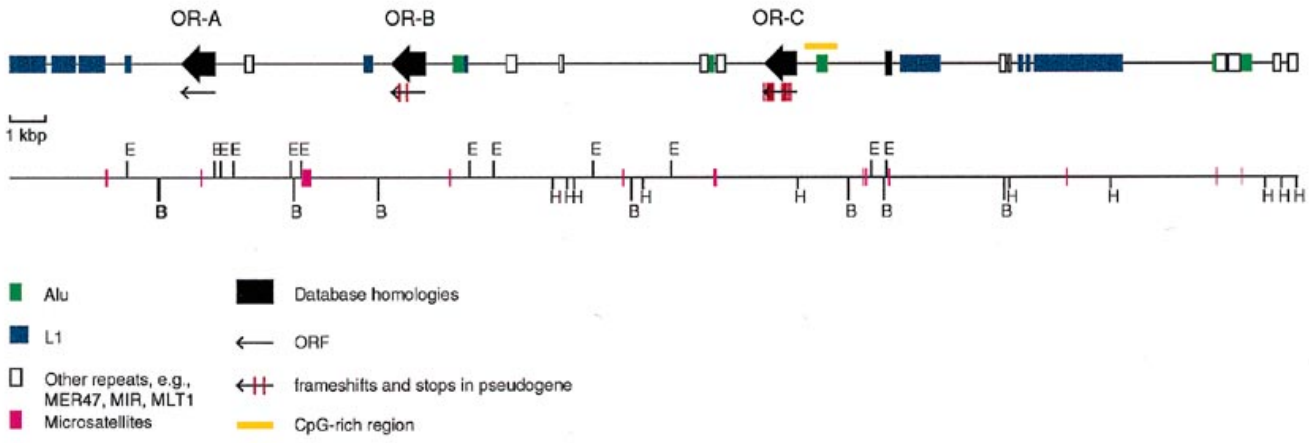


Figure 5. Prominent features of cosmid f7501's 36 266-bp sequence. Regions of significant homology to olfactory receptor (OR) genes (large black arrows) and interspersed repeats (Alu, L1, etc.) are shown at the top. The predicted and confirmed restriction map is shown at the bottom. The small black rectangle indicates the almost identical match to the vector-proximate sequence of half-YAC yrm2001a (27) (accession no. U11826) (see text).

repeatedly tested positive for several PCR assays outside the region of simple overlap (Fig. 3b). f7501 has no homology to any other previously described subtelomeric repeat.

No internal repeats, a feature of many subtelomeric clones, are found within the cosmid. Cosmid f7501 is relatively rich in L1 interspersed repeats (occupying 18% of nucleotides). Assorted other interspersed repeats comprise an additional 9.5% of the sequence. A variety of microsatellites, including a 260-bp imperfect TA repeat, are present, as indicated in Figure 5.

Humans vary in the number and location of these OR sequences

Although variation in the chromosomal location and copy number of the putative OR gene was suspected on the basis of the behavior of the entire cosmid, we wished to demonstrate polymorphism involving the OR sequences directly. Three approaches were used. First, small subclones containing each of the OR sequences were FISH mapped to five individuals. Second, various chromosomes were flow-sorted from each of three individuals into separate spots on nitrocellulose filters (seven to 12 chromosomes/individual; 10^4 chromosomes/spot) and probed for the OR-7501A sequence by hybridization. In both of these experiments, the OR sequences showed a polymorphic chromosomal distribution concordant with the FISH results of the entire cosmid (data not shown). In the third experiment, specific chromosomes were isolated from two individuals by flow sorting and assayed by PCR for the OR-7501A sequence. The OR-7501A sequence amplified chromosomes 15 and 19 in both individuals (chromosome 3 was not tested), but it was polymorphically present or absent on chromosomes 7, 11 and 16 (Fig. 7). Again, the polymorphic distribution of the OR sequence was the same as that of the entire cosmid.

DISCUSSION

Our analyses of large blocks of DNA that are duplicated in the subtelomeric regions of multiple chromosomes provide insight into the manner in which variation and complexity of the human genome is generated. The 36-kb segment identified by cosmid f7501, which we have studied in detail, serves as an example of a larger collection of duplicated, subtelomeric sequence blocks. This segment distinguishes itself from previously reported subtelomeric segments in several respects. It contains three members of the olfactory receptor gene family, one of which is potentially functional. The segment is extraordinarily polymorphic: in our survey of chromosomes from 45 unrelated individuals, the segment could be found at 14 different locations, although f7501-positive alleles were particularly rare at some termini (2q, 6p, 6q, 8p, and 19q). As a consequence of this polymorphic distribution, normal individuals carry seven to 11 copies of this OR-containing block in their genomes. The distribution pattern of f7501 varies among human populations. Despite its presence on virtually every human homolog of 3q, 15q, and 19p analyzed from different human populations, only a single, nonorthologous site contains the bulk of the sequence in chimpanzee and gorilla. Paralogous copies of this segment are very similar in structural detail. The full extent of the homology among chromosomes is unknown, but exceeds ~90 kb. The structure of the region is complex: sequences flanking the OR-containing segment are duplicated on different sets of chromosomes, both larger than the set containing the f7501 segment itself.

These structural and distributional characteristics demonstrate the dynamic nature of the subtelomeric regions. Inter-chromosomal exchanges of large DNA sequence blocks appear to have modified these regions relatively recently. The presence of OR sequences in these rearrangement-prone regions suggests that the subtelomeric zones may serve as an evolutionary test-bed for this multigene family. These issues are discussed in more detail in the following sections.

Subtelomeres are a patchwork of blocks duplicated on different subsets of chromosomes

Duplications near telomeres are wide-spread and frequent. The emerging picture of the subtelomeric zones is one of a mosaic of many different segments, each duplicated on a characteristic subset of chromosomes. Of 20 clones that have been characterized from subtelomeric regions (our unpublished results and refs

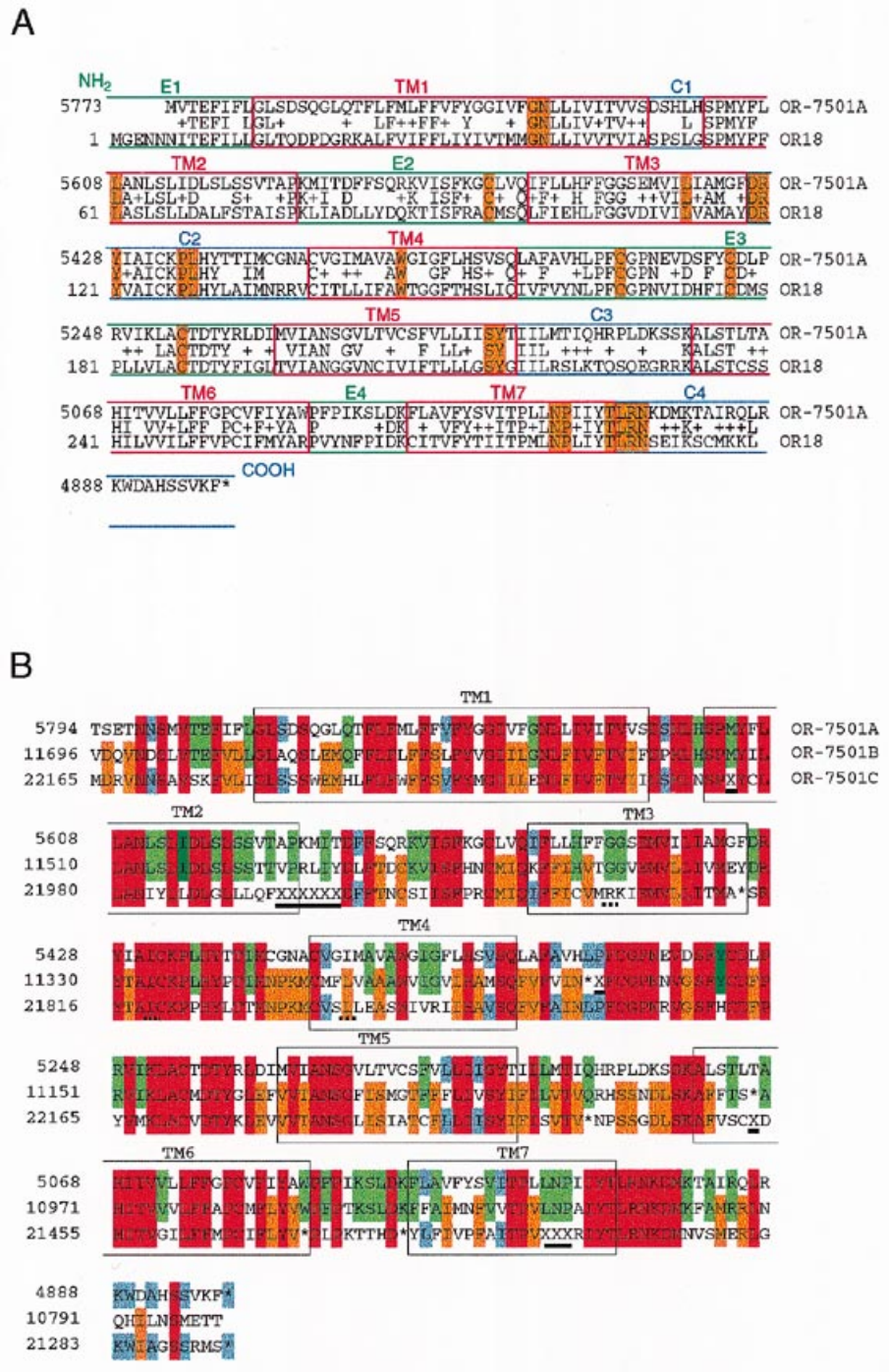


Figure 6. (**A**) Alignment of the predicted protein of OR-7501A and OR18, a rat olfactory gene (S29710) identified from mRNA from olfactory neuroepithelium (34). The two have 49% amino-acid identity (144/294 residues) and 70% similar amino-acids (205/294). The BLAST score was 782, and $p = 1.9 \times 10^{-103}$. The transmembrane (TM), extracellular (E) and cytoplasmic (C) domains are indicated. The residues highlighted in yellow have been described as conserved and potentially essential (33). The numbering of the OR-7501A sequence refers to nucleotide position in the GenBank submission of f7501's sequence (L78442). The numbering of OR18 refers to its amino-acid sequence. (**B**) Amino-acid alignment of the three OR-like sequences in f7501. The numbering corresponds to the nucleotide position in the GenBank submission of f7501's sequence (L78442). Dark lines indicate the locations of frame-shift causing deletions in OR-7501B and OR-7501C (of 1 nt before nt 11197 in OR-7501B, 1 nt before nt 21992 and 21461, 17 nt before nt 21935, and 8 nt before nt 21338 in OR-7501C). Dashed lines indicate the positions of single nucleotide insertions causing frame-shifts in OR-7501C. These frameshifts were corrected to produce the amino-acid alignments. Color is used to indicate amino-acid identities among the three sequences.

7,8,16–30), each has a different, usually polymorphic, chromosomal distribution. Some sequences, like f7501, are shared by relatively few termini in a given individual. Others have a much wider distribution [e.g. cosmid 16432 contains sequences shared by at least 16 termini and nine interstitial sites (30)]. The subtelomeres of all but five of the 43 tips (disregarding the short

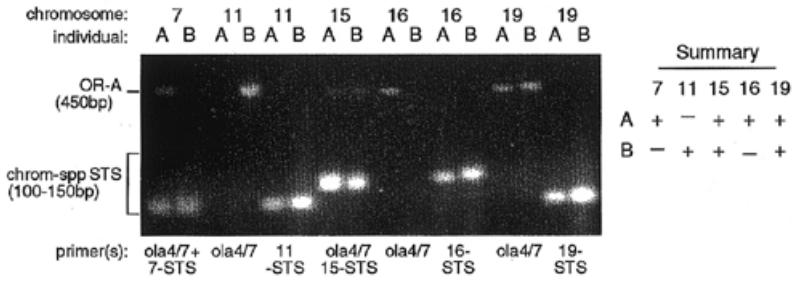


Figure 7. PCR assays on flow-sorted chromosomes show that OR-7501A is present on different chromosomes in different individuals. Flow-sorted chromosomes were subjected to PCR with primers OLA7 and OLA4, which lie within and 5′ of the putative coding region, respectively. Product of the expected size was amplified from chromosomes 7, 15, 16, and 19 sorted from GM10473 (individual A), and from chromosomes 11, 15, 19 sorted from GM10967 (individual B), in perfect concordance with the FISH results (Figs 2 and 1B). Chromosome 3 was not tested. An STS specific for each chromosome was included as a positive control, either in the same tube (chromosomes 7 and 15) or in a replicate tube (chromosomes 11, 16, and 19).

arms of the five acrocentric chromosomes) contain sequences common to one or more other chromosomes. The map of 3q, for example, is linked to the maps of 38 other chromosome arms through a variety of shared subtelomeric sequences.

The overall size of these subtelomeric zones has not been established yet. A minimal estimate can be inferred by the fact that unique, chromosome-specific sequences have been isolated from the proximal ends of half-YACs that extend ~200–300 kb from the telomere (38). However, this estimate does not take large-scale variation into account, and it assumes that zones of duplicated blocks do not lie proximal to these unique segments.

Our results indicate that the transitional zone from telomere to unique sequence DNA has a more complex structure than may have been suspected previously and confirm the recent findings of Flint and coworkers (15). The homology among chromosomes does not decrease monotonically from the telomere to single-copy DNA. Rather, segments shared by only a few chromosomal ends are flanked by segments shared by many ends. Our PCR analyses of the monochromosomal panel (Fig. 3) and FISH analyses of overlapping P1 clones (Fig. 4) support this conclusion. Chromosome 6 is particularly instructive; it contains sequences situated proximal and distal of f7501 on other chromosomes, but lacks the majority of f7501. The processes that have produced this irregular patchwork have also generated large-scale variation.

The potential impact of large-scale polymorphisms on meiotic pairing

The meiotic process must accommodate the significant disparity between the tips of homologous chromosomes and extensive similarity among heterologous chromosomes. This point was raised earlier by Wilkie *et al.* (7), who identified common alleles of 16p that differed by 200 kb in length. Although the full extent of the variation at each telomere is not known, we show here that alleles of as many as 13 chromosomes differ by the presence or absence of the 36-kb f7501 block. We have identified alleles of 9q and 11p that differ by at least 85 kb in length (Fig. 4). At the same time, the tips of heterologous chromosomes can share >90 kb of homologous sequence. Given that meiotic pairing of homologous chromosomes initiates near telomeres (3), we hypothesize that pairing is occasionally disturbed by this confusing situation. Indeed, abnormal pairing of telomeres of heterologous chromosomes is observed with relatively high frequency in female meioses (39). The frequency with which these mispairings lead to deleterious rearrangements is not known. However, subtelomeric rearrangements are increasingly being recognized in association with congenital anomalies (40), as the reagents for tracking the most distal unique portions of chromosomes improve (38,41,42).

The evolutionary history of this subtelomeric duplication

Our results provide information on the evolutionary history of the subtelomeric zone. The dissimilarity of the three OR sequences within the f7501 block (only 62–68% identical nucleotides) implies that they arose by duplication over 40–100 million years ago, if they diverged at the rate observed for orthologous sequences [$1–4 \times 10^{-9}$ mutations per site per year (43)] or globin paralogs [13×10^{-9} (44)].

This group of OR sequences has since been duplicated on additional chromosomes. The population analyses provide clues to the times of these events. Because virtually all copies of chromosomes 3q, 15q, and 19p sampled from different populations contain the f7501 segment, its duplication onto these three locations must predate the spread of humans around the world. The dispersed, multi-copy pattern in humans is unique among the primates we analyzed, indicating that the OR-containing segment has changed copy number and location during primate evolution. Both chimpanzee and gorilla carry the bulk of the f7501 sequence at a single location, which is not orthologous to any of the human sites. The sequence therefore marks a rearrangement, which was not previously recognized, that distinguishes the primate orthologs of human chromosome 4. The multiplicity of copies in the human genome might be explained by a burst of transpositional activity in the human genome since the trichotomy of the human/chimpanzee/gorilla-clade [as seen with pericentromeric copies of sequence from the telomere of Xq28 (45)]. In this scenario, the ancestral site might have been syntenic with human 15q, where the sequence is retained in orangutan and human. However, its transfer to 4q in both chimpanzee (PTR3q) and gorilla (GGO3q) is difficult to reconcile with recent molecular data that place chimpanzees closer to humans in the clade (46). An alternative scenario is suggested by the appearance of very small signals, presumably representing a small portion of f7501's sequence, on the chimpanzee orthologs of the four most common human sites. A larger block may have been present on many ancestral chromosomes (e.g. the ancestral orthologs of HSA3q,

4q, 11p, 15q, and 19p) and the bulk of the f7501 sequence was lost from all but 4q in chimpanzee and gorilla, from all but 15q in orangutan, and from only 4q in humans. Our preliminary data support such a complex history: sequences flanking the f7501 block hybridize to multiple telomeres in chimpanzee (C. Friedman and B.Trask, unpublished).

The fluidity of the subtelomeric regions is also evident from the polymorphic and more infrequent appearance of the f7501 block on 11 other human chromosomes. Either these sites gained or lost f7501 relatively recently, or they represent old duplications that were not fixed in all the ancestors of modern man. For example, the f7501 block may have transferred only recently to the tip of chromosome 7p and spread within the Pygmy populations, where it is found at high frequency. Alternatively, the f7501 sequence may have been common at 7p among ancestors of humans, but its frequency could have been reduced through random drift or selective loss in groups that migrated out of Africa. The 7p copy may have been present in the founding Pygmy population and become fixed by random drift. The most likely scenario should become obvious from future sequence comparisons of the blocks on different chromosomes. Regardless of their origin, the significant population differences we observe suggest that the presence or absence of subtelomeric blocks on particular chromosomes could serve as additional markers with which to study the phylogeny of modern man.

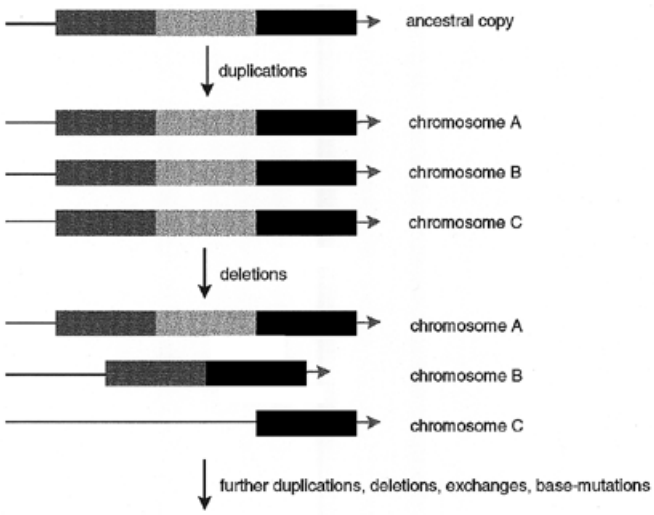


Figure 8. Working model of events that may have led to the current structure and variability of subtelomeric regions.

A model for the evolution of subtelomeric zones

Taking all of these results into account, we postulate that some chromosomes lost portions of larger ancestral duplications as shown in Figure 8. In this working model, large blocks of sequence translocated periodically from one chromosome to another. The number of genomic copies could have increased if block-containing alleles of both donor and recipient chromosomes were passed on to subsequent generations. Such interchromosomal exchange has been postulated before to explain the multichromosomal distribution of other subtelomeric sequences (7,17,29). The structure we observe around the f7501 segment suggests that subsequent interchromosomal exchanges further modified the organization of these blocks. Unequal exchanges could result in apparent interstitial insertions or deletions in the exchange partners. The extensive homology among different telomeres could have facilitated these exchanges. The addition of subtelomeric blocks to new sites or their loss from existing sites appear from our population study to have occurred infrequently during human evolution, relative to the rate with which particular variants accumulated by random drift.

The remarkable structural similarity of blocks on different chromosomes suggests that interlocus exchange between already homologous tips may occur frequently. For example, the similarity of chromosome 3 and chromosome 19 is striking in light of our population analyses, which date the duplication of the f7501 block onto these chromosomes before the time of the last common ancestor of all humans, and perhaps before our divergence from other primates. The tips of 3q and 19p show homology over at least 85 kb of sequence. We identified alleles of these two chromosomes whose restriction maps with three different enzymes are identical over at least 34 kb. Preliminary sequence comparisons indicate that these paralogs are more similar than are two alleles of 19 (B. Trask *et al.*, unpublished data). This similarity suggests that interchromosomal exchanges homogenized sequences on different chromosomes after the initial duplications occurred. Indeed, interchromosomal transfer between homologous sequences near the telomeres of 4q and 10p was observed directly in a recent report (47).

Are the subtelomeric regions nurseries for OR-family diversity?

The discovery of OR genes in these rearrangement-prone regions raises the possibility that subtelomeres might function as a nursery for the generation of diversity in this multigene family. The OR gene family has 500–1000 members (48) that range in similarity from 45 to 98% (33,49). Our results provide a possible explanation for how part of this large, diverse family might have emerged.

We hypothesize that subtelomeres serve as a place where OR sequences can be duplicated and modified, without affecting more proximal dosage-sensitive genes. Association of a subset of the OR family with these dynamic regions may be advantageous in adapting to new environments. Precedent for this mechanism can be found in more primitive organisms, such as *Giardia* (50) and *Plasmodium* (51). The antigen-encoding genes of *P.falciparum* are found in subtelomeric regions, and exchanges of subtelomeric sequences among chromosomes generate marked length-polymorphism and antigen diversity (51). Corcoran *et al.* (52) speculated that 'it may be advantageous [for these genes] to be near the dynamic recombining sequences of the telomeres.' The same may be true of OR genes.

In accordance with this hypothesis, randomly selected OR genes show a strong bias for terminal regions in the human genome (ref. 49 and unpublished results). It is also notable that six of the 11 known interstitial locations of expressed murine OR genes (53) are syntenic with telomeric regions of human chromosomes. Inversions or fusions may have internalized a subset of OR genes during the evolution of mouse chromosomes. Such a process was also postulated to explain the location of IL9 and HBA genes near the telomere of human 16p and at an interstitial site on mouse chromosome 11 (29,54).

The telomeres house members of other gene families, such as homeobox domains, helicases, and interleukin-receptors (26,28,55). Subtelomeric duplications and rearrangements may also play a role in spawning new functionalities in these families. Of course, many of the sequences generated in subtelomeric regions will be functional failures. Two of the three OR genes in f7501 and many of the subtelomeric copies of these other gene families are pseudogenes (ibid). Pseudogene copies of a large number of expressed sequences (ESTs) are also found in the subtelomeric region of 16p (14).

The possibility that sequences varying in copy number and position among humans could contain functional genes has not been considered before. If subtelomeric OR genes are expressed, the variation in OR copy number we report here offers a possible explanation for observed polymorphism in thresholds to various odorants (48).

Concluding remarks and implications for mapping the human genome

Our observations raise further questions about the generation and nature of genomic diversity. Are subtelomeres nurseries of functional diversity, or are they merely genetic junkpiles of phylogenetic markers? How large and variable are the subtelomeric regions, and what influence do they have on meiotic pairing? Do subtelomeric duplications on different chromosomes evolve independently or do they recombine to blur their phylogenetic history? Answers to these questions will come only through further mapping and sequencing of paralogous subtelomeric regions.

The size, similarity and polymorphic nature of these duplications pose challenges to mapping methods honed on single-copy regions of the genome. Our findings therefore have practical consequences relevant to efforts to map the human genome. Contigs assembled from clones derived from large, highly similar paralogs will be branched. Many clones will be classified during contig-construction as deleted or chimeric, yet they represent bona fide segments of one or more chromosomes. If maps are constructed using DNA from individuals who happen to carry short variants, large gene-containing segments may be inadvertently omitted. For example, 284 kb of contiguous sequence spanning from the telomere to unique sequence on 16p provides detailed insight into the structure and function of subtelomeres (14), yet it is the description of a short 16p allele, which apparently lacks the OR-containing segment we describe here. Strategies that employ chromosome-specific reagents from a large number of individuals will be needed to unravel these duplications, sort out their phylogeny, and describe this form of genetic diversity. Our findings demonstrate that such strategies will contribute to our understanding of the evolution of primates, chromosomes, and multigene families.

MATERIALS AND METHODS

Cell lines

EBV-transformed lymphoblast cell lines established of 35 humans from seven populations were obtained from NIGMS repository and are identified by the prefix GM. Primary PHA-stimulated peripheral blood cultures were analyzed for an additional eight anonymous donors after informed consent. CGM1, the lymphoblast cell line used to prepare the Washington University half-YAC library (56), and 978SK, the skin fibroblast line used to prepare the Research Genetics CITB-978SK-B BAC library (57), were also analyzed. Peripheral blood samples of three common chimpanzees and one gorilla were generously provided by the San Francisco and San Diego Zoos and subjected to short-term culture in the presence of PHA and phorbol (58). Other primate cell lines (chimpanzee skin fibroblast GM03542, chimpanzee lymphoblast line TANK, gorilla lymphoblast line CRL1854/ROK, and orangutan skin fibroblast GM06213) were obtained from either NIGMS and ATCC. Monochromosomal hybrids for chromosomes 3 (GM10253), 15 (GM11418, referred to as 15A in figures, and GM11715, 15B), 19 (GM10449, 19A, and GM10612, 19B), and 11 (GM10927A, GM11087, and GM11941) were used for FISH, Southern blot and PCR analyses.

Clones

f7501 was derived from a cosmid library constructed in Lawrist 5 using DNA enriched for human chromosome 19 by flow-sorting from monochromosomal hybrid GM10449 (HL9–5B) (30,59). The DNA was digested partially with *Mbo*I and cloned into the *Bam*HI site of the vector. Subclones were prepared by sonication of the cosmid and cloned in Bluescript SK+. OR-containing subclones were identified through sequence analysis and confirmed by PCR. *Eco*RI fragments purified by electrophoresis in LMP agarose were also subcloned in Bluescript. P1 clones RMC0MP001 and RMC0MP012-015 were isolated by PCR screening of the DuPont P1 genomic library with primers OLA4/OLA7 (pair 6′, see below). The coordinates of these P1s are 122h8, 1137c1, 697h9, 1427b3, and 227a10, respectively. P1 RMC0MP013 was also identified as the only positive clone in a screen with primer pair 18 (see below). Cosmids 176C1 and 214B2 were isolated by PCR screening with OLA4/OLA7 primers from a chromosome 3-specific library, constructed from chromosomes flow-sorted from monochromosomal hybrid GM10253. The chromosome 19 and chromosome 3 cosmids came from different chromosomes, isolated in different hybrids, using DNA from different individuals (60).

Fluorescence *in situ* hybridization (FISH)

Cells were cultured, synchronized by a methotrexate block, released in the presence of BrdU, and harvested to prepare metaphase spreads, according to published procedures (61). DNA from cosmid f7501, plasmid subclones, or P1 clones was biotinylated by nick translation and hybridized to metaphase cells fixed on slides. Methods for preparation of the slides and probe, hybridization, washing, detection with FITC, fluorescent banding, and analysis are described elsewhere (61). In most experiments with f7501, the chromosomes were QFH-banded with DAPI and actinomycin and counterstained with PI. In this case, two photographs were collected on color slide film. By marking clear-overlays, the sites of FITC signals, superimposed on the PI-counterstain in one photograph, were mapped relative to the DAPI bands in the second photograph. In later experiments, the FITC and DAPI were collected in registration and displayed in false colors using a Princeton cooled CCD camera, Chroma-Technology spectral filters, and image-analysis software (IP Lab Spectrum). In several cases, the probes were mapped in a single image with respect to a fluorescent R-banding pattern produced by propidium iodide staining as described (61). Primate chromosomes were identified through a combination of DAPI banding

and 'painting' with a collection of probes derived from specific human chromosomes. Paints were produced by flow-sorting 100 chromosomes from a karyotypically normal lymphoblast cell line and subjecting this material to universal amplification (DOP-PCR) in the presence of digoxigenin-labeled nucleotides (61,62). The painting probe was hybridized along with biotinylated f7501, and detected with Texas-Red labeled antibodies (61). FITC, Texas-Red and DAPI counterstain were analyzed simultaneously by viewing metaphases through dual and triple bandpass filters (ChromaTechnology). In each experiment, the locations of hybridization sites were scored in a minimum of 10 metaphases per probe per individual. When heterozygosity is indicated (gray symbols) in Figures 2 or 4, signals were present on one, and never two, homologs per cell.

Sequencing

Cosmid f7501 was sequenced through a combination of random-shotgun and directed strategies. Shotgun clones were initially generated by sonicating the cosmid and cloning in Bluescript. Both ends of each clone were sequenced using dye-primer chemistry on a 373 DNA Sequencer (Applied Biosystems). Additional sequence was derived from the ends of nested fragments generated by exonuclease digestion of cloned *Eco*RI fragments. In this first phase, sequence reads averaging 340 bp, in theory representing 3.8-fold coverage, were assembled into 26 contigs. Due to significant *Escherichia coli* contamination, a second shotgun library was prepared in M13. An additional 6-fold coverage was obtained by sequencing, using dye-primer chemistry, randomly selected clones from this library and from a set derived from the 11.9 kb *Eco*RI subclone, which was added to the assembly to compensate for the low representation of this region in the first phase. The sequence was assembled using DNAStar. Gaps were closed with reverse reads generated from PCR products of selected inserts of M13 clones. Ambiguities were resolved by sequencing selected M13 clones with the dye terminator chemistry. End sequences were obtained from P1s by direct T7- and Sp6-primed sequencing of Qiagen-purified DNA as directed in the ABI Prism dye terminator cycle sequencing kit (Perkin Elmer), except that 2 μg DNA and 50 pmol primer were used per reaction.

Sequence analysis

The final sequence was analyzed using RepeatMasker (63) to screen for a variety of interspersed repeats and sputnik (64) for microsatellite repeats. Both programs are based on cross_match (65). The sequence was compared with published nucleotide and protein sequences by BlastN and BlastX. Alignments of OR sequences were performed and percentage similarities calculated using both cross_match and PILEUP in the GCG package. In order to calculate the amino-acid homology of OR-7501B and OR-7501C with other OR genes, their frameshifts were 'corrected' at the nucleotide level and scored as mismatches in the predicted protein sequence. The annotated sequence has been submitted to public databases under accession number L78442.

Restriction mapping

The restriction map of f7501 predicted from its sequence was confirmed by sizing fragments produced by complete single and double digestion with *Eco*RI, *Bam*HI, and *Hin*dIII on 1% agarose gels. In addition, the cosmid was digested to completion with *Sfi*I, which cuts at sites flanking the cloning site, and then subjected to partial digest mapping with *Eco*RI. The partial digests were electrophoresed on 0.7% agarose gels and probed with ^{32}P-labeled Sp6 or T7 oligomers flanking the insert. The *Eco*RI, *Bam*HI and *Hin*dIII restriction digests of cosmids 176C1, 214B2, and f7501 were compared after electrophoresing the fragments on a 19 cm 1% agarose gel run 16 h at 2 V/cm at 16°C, staining with 1:20 000 dilution of SYBR-green I (Molecular Probes), and scanning on a FluorImager 575 (Molecular Dynamics) as described (66). The approximate size of the P1 clones was determined by summing the lengths of *Eco*RI restriction fragments.

Southern blot analyses

*Eco*R1 digested genomic DNA from several human cell lines (CGM1, GM10473, GM10977, GM11525), shown by FISH to carry seven to 11 copies of the f7501 sequence, and from monochromosomal hybrid lines for chromosomes 3, 15 and 19 (see cell lines) were blotted onto nylon membranes and hybridized with ^{32}P labeled probes following conventional procedures (67). The probes were random-prime labeled subclones or PCR products recognizing the 2.2, 2.5, 2.8*, 4.7*, 5.6, and 11.9 kb *Eco*RI fragments of f7501 (*: RFLP detected among paralogs). The final washes of the blots were in 0.1× SSC at 65°C. Probes for OR-7501A, -B, and -C do not cross-react at this stringency.

PCR primers and typing of hybrid panels

Twenty-four PCR assays were designed using primers defined from the sequence of f7501. These primer pairs are identified by number. PCR assay 6′, used in library screens and to type flow-sorted chromosomes, employed primers OLA7 (5′ CTC ACC GAA TGG AGA AAG CC, corresponding to bases 5322–5341 in f7501's sequence) and OLA4 (5′ TCT GAC TTC CTT CTC CTT CTC, corresponding to the reverse complement of bases 5837–5857) at an annealing temperature of 62°C. PCR assay 18, also used to screen the P1 library, employed primers 10550a (5′ TTA CAG CCT GAC AGG GCC, corresponding to nt 33283–33300) and 10550b (5′ GGT CTA GAG CTG CTC TAG G, corresponding to the reverse complement of bases 33703–33721) at 60°C. The other primer sequences, their position in the f7501 sequence, and annealing temperatures are available upon email request (trask@biotech.washington.edu). PCR reactions (25 μl) were conducted following conventional procedures using 100 ng of DNA from the NIGMS monochromosomal hybrid panel #2, several additional monochromosomal hybrids (see cell lines), or human genomic DNA, and 1 ng of DNA from cosmid, P1 and half-YAC clones. The products were analyzed on ethidium-bromide stained 1% agarose gels.

PCR analyses of flow-sorted chromosomes

Chromosomes were isolated from lymphoblast cell lines GM10473 and GM10967 into a polyamine buffer, stained with Hoechst 33258 and chromomycin A3, and flow-sorted using a custom dual-laser flow cytometer as described elsewhere (68). One thousand copies of each of five chromosomes were sorted from each individual directly into 5 μl of sterile water in 0.5-ml PCR tubes and frozen at –20°C before use. PCR buffer (25 mM

TAPS, 16.6 mM $(NH4)_2SO_4$, 2.5 mM $MgCl_2$, 0.5% BSA, and 0.5 mM β-mercaptoethanol), dNTPs (200 μM each), and primers (1 μM) were added, the chromosomes were denatured at 94°C for 10 min before 1.25 U of each AmpliTAQ (Perkin-Elmer) and Pwo polymerase (Boehringer Mannheim) were added, and the 25 μl reactions were subjected to 35 amplification cycles of 94°C 10 s, 62°C 30 s, and 72°C 60 s (as per N. Carter, personal communication). Primers OLA4/OLA7 (see above) were used to amplify the OR-7501A region. These primers amplify only from chromosomes 3, 15, and 19 in the hybrid panel (assay 6′ in Fig. 3) and do not amplify the OR-7501B or OR-7501C sequences. The sorts were also tested for amplification with chromosome-specific STS primers [1458, 1263, 2417, 1232, and 1350 for chromosomes 7, 11, 15, 16, and 19, respectively (69)], separately or in the same reaction as the OLA primers. The products were analyzed on ethidium bromide-stained 1% agarose gels. Water controls were negative.

Hybridization to chromosomes flow-sorted onto filters

Ten thousand copies of each of a variety of chromosomes were flow-sorted from three different individuals (GM10469, GM10473, GM10977) into small spots on Hybond+ filters. The chromosomes were denatured and affixed to the membranes as described (70). Filters were washed to a stringency of 0.1× SSC at 65°C after hybridization with a probe amplified with OLA4/OLA7 primers from F7501 and random-prime-labeled with ^{32}P.

ACKNOWLEDGEMENTS

We thank Mary Ellen Ahearn and Jane Lamerdin for assistance in sequencing, Heather Mefford and Leah LaTray for flow-sorting support, Harold Riethman for half-YAC clones, and the San Francisco and San Diego Zoos for primate blood samples. This research was supported by NIH grants HG00256, HG00470, and CA576626, DOE grants DE-FG06-93ER61553, DE-FG06-93ER61662, DE-FG03-96ER62173, and DE-FG06-92ER61467, and the UCSF Program for Analytical Cytology. Work by A.M.-G. was performed at LLNL under U.S.DOE contract W-7405-ENG-48.

REFERENCES

1. Saccone, S., De-Sario, A., Wiegant, J., Raap, A.K, Della-Valle, G. and Bernardi, G. (1993) Correlations between isochores and chromosomal bands in the human genome. *Proc. Natl Acad. Sci. USA* **90**, 11929–11933.
2. Laurie, D.A. and Hultén, M.A. (1985) Further studies on bivalent chiasma frequency in human males with normal karyotypes. *Ann. Hum. Genet.* **49**, 189–201.
3. Barlow, A.L. and Hultén, M.A. (1996) Combined immunocytogenetic and molecular cytogenetic analysis of meiosis I human spermatocytes. *Chromo. Res.* **4**, 562–573.
4. Zakian, V.A. (1995) Telomeres: Beginning to understand the end. *Science* **27**, 1601–1607.
5. Harley, C.B., Futcher, A.B. and Greider, C.W. (1990) Telomeres shorten during ageing of human fibroblasts. *Nature* **345**, 458–460.
6. Hastie, N.D., Dempster, M., Dunlop, M.G., Thompson, A.M., Green, D.K. and Allshire, R.C. (1990) Telomere reduction in human colorectal carcinoma and with ageing. *Nature* **345**, 866–868.
7. Wilkie, A.O., Higgs, D.R., Rack, K.A., Buckle, V.J., Spurr, N.K., Fischel-Ghodsian, N., Ceccherini, I., Brown, W.R. 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.
8. Youngman, S., Bates, G.P., Williams, S., McClatchey, A.I., Baxendale, S., Sedlacek, Z., Altherr, M., Wasmuth, J.J., MacDonald, M.E., Gusella, J.F., Sheer, D. and Lehrach, H. (1992) The telomeric 60 kbp of chromosome arm 4p is homologous to telomeric regions on 13p, 15p, 21p, and 22p. *Genomics* **14**, 350–356.
9. Riethman, H.C., Spais, C., Buckingham, J., Grady, D. and Moyzis, R.K. (1993) Physical analysis of the terminal 240 kb of DNA from human chromosome 7q. *Genomics* **16**, 25–32.
10. Cook, G.P., Tomlinson, I.M., Walter, G., Riethman, H., Carter, N.P., Buluwela, L., Winter, G. and Rabbitts, T.H. (1994) A map of the human immunoglobulin VH locus completed by analysis of the telomeric region of chromosome 14q. *Nature Genet.* **7**, 162–168.
11. Macina, R.A., Negorev, D.G., Spais, C., Ruthig, L.A., Hu, X.L. and Riethman, H.C. (1994) Sequence organization of the human chromosome 2q telomere. *Hum. Mol. Genet.* **3**, 1847–1853.
12. Negorev, D.G., Macina, R.A., Spais, C., Ruthig, L.A., Hu, X.L. and Riethman, H.C. (1994) Physical analysis of the terminal 270 kb of DNA from human chromosome 1q. *Genomics* **22**, 569–578.
13. Reston, J.T., Hu, X.L., Machina, R.A., Spais, C. and Riethman, H.C. (1995) Structure of the terminal 300 kb of DNA from human chromosome 21q. Genomics **26**, 31–38.
14. Flint, J., Thomas, K., Micklem, G., Raynham, H., Clark, K., Doggett, N.A., King, A. and Higgs, D.R. (1997) The relationship between chromosome structure and function at a human telomeric region. *Nature Genet.* **15**, 252–257.
15. Flint, J., Bates, G.P., Clark, K., Dorman, A., Willingham, D., Roe, B.A., Micklem, G., Higgs, D.R. and Louis, E.J. (1997) Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Hum. Mol. Genet.* **6**, 1305–1314.
16. Cheng, J.F., Smith, C.L. and Cantor, C.R. (1989) Isolation and characterization of a human telomere. *Nucleic Acids Res.* **17**, 6109–6127.
17. Brown, W.R.A., MacKinnon, P.J., Villasante, A., Spurr, N., Buckle, V.J. and Dobson, M.J. (1990) Structure and polymorphisms of human telomere-associated DNA. *Cell* **63**, 119–132.
18. Cross, S., Lindsey, J., Fantes, J., McKay, S., McGill, N. and Cooke, H. (1990) The structure of a subterminal repeated sequence present on many human chromosomes. *Nucleic Acids Res.* **18**, 6649–6657.
19. de Lange, T., Shiue, L., Myers, R.M., Cox, D.R., Naylor, S.L., Killery, A.M. and Varmus, H.E. (1990) Structure and variability of human chromosome ends. *Mol. Cell Biol.* **10**, 518–527.
20. Weber, B., Collins, C., Robbins, C., Magenis, R.E., Delaney, A.D., Gray, J.W. and Hayden, M.R. (1990) Characterization and organization of DNA sequences adjacent to the human telomere associated repeat TTAGGGn. *Nucleic Acids Res.* **18**, 3353–3361.
21. Wells, R.A., Germino, G.G., Krishna, S., Buckle, V.J. and Reeders, S.T. (1990) Telomere-related sequences at interstitial sites in the human genome. *Genomics* **8**, 699–704.
22. Ijdo, J.W., Baldini, A., Ward, D.C., Reeders, S.T. and Wells, R.A. (1991) Origin of human chromosome 2: An ancestral telomere-telomere fusion. *Proc. Natl Acad. Sci. USA* **88**, 9051–9055.
23. Ijdo, J.W., Lindsay, E.A., Wells, R.A. and Baldini, A. (1992) Multiple variants in subtelomeric regions of normal karyotypes. *Genomics* **14**, 1019–1025.
24. Weber, B., Allen, L., Magenis, R.E. and Hayden, M.R. (1991) A low-copy repeat located in subtelomeric regions of 14 different human chromosomal termini. *Cytogenet. Cell. Genet.* **57**, 179–183.
25. Hoglund, M., Mitelman, F. and Mandahl, N. (1995) A human 12p-derived cosmid hybridizing to subsets of human and chimpanzee telomeres. *Cytogenet. Cell Genet.* **70**, 88–91.
26. Kermouni, A., van Roost, E., Arden, K.C., Vermeesch, J.R., Weiss, S., Godelaine, D., Flint, J., Lurquin, C., Szikora, J.-P., Higgs, D.R., Marynen, P. and Renauld, J.-C. (1995) The IL-9 receptor gene (IL9R): Genomic structure, chromosomal localization in the pseudoautosomal region of the long arm of the sex chromosomes, and identification of IL9R pseudogenes at 9qter, 19pter, 16pter, and 18pter. *Genomics* **29**, 371–382.
27. Martin-Gallardo, A., Lamerdin, J., Sopapan, P., Friedman, C., Fertitta, A., Garcia, E., Carrano, A., Negorev, D., Macina, R.A., Trask, B.J. and Riethman, H.C. (1995) Molecular analysis of a novel subtelomeric repeat with polymorphic chromosomal distribution. *Cytogenet. Cell Genet.* **71**, 289–295.
28. Winokur, S.T., Bengtsson, U., Vargas, J.C., Wasmuth, J.J. and Altherr, M.R. (1996) The evolutionary distribution and structural organization of the homeobox-containing repeat D4Z4 indicates a functional role for the ancestral copy in the FSHD region. *Hum. Mol. Genet.* **5**, 1567–1575.

29. Vermeesch, J.R., Petit, P., Kermouni, A., Renauld, J.-C., van den Berghe, H. and Marynen, P. (1997) The IL-9 receptor gene, located in the Xq/Yq pseudoautosomal region, has an autosomal origin, escapes X inactivation and is expressed from the Y. *Hum. Mol. Genet.* **6**, 1–8.
30. Trask, B., Fertitta, A., Christensen, M., Youngblom, J., Bergmann, A., Copeland, A., de Jong, P., Mohrenweiser, H., Olsen, A., Carrano, A. and Tynan, K. (1993) Fluorescence in situ hybridization mapping of chromosome 19: cytogenetic band location of 540 cosmids and 70 genes or DNA markers. *Genomics* **15**, 133–145.
31. Jauch, A., Wienberg, J., Stanyon, R., Arnold, N., Tofanelli, S., Ishida, T. and Cremer, T. (1992) Reconstruction of genomic rearrangements in great apes and gibbons by chromosome painting. *Proc. Natl Acad. Sci. USA* **89**, 8611–8615.
32. Yunis, J.J. and Prakash, O. (1982) The origin of man: a chromosomal pictorial legacy. *Science* **215**, 1525–1530.
33. Buck, L. and Axel, R. (1991) A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell* **65**, 175–187.
34. Raming, K., Krieger, J., Strotmann, J., Boekhoff, I., Kubick, S., Baumstark, C. and Breer, H. (1993) Cloning and expression of odorant receptors. *Nature* **361**, 353–356.
35. Issel-Tarver, L. and Rine, J. (1996) Organization and expression of canine olfactory receptor genes. *Proc. Natl Acad. Sci. USA* **93**, 10897–10902.
36. Parmentier, M., Libert, F., Schurmans, S., Schiffmann, S., Lefort, A., Eggerickx, D., Ledent, C., Mollereau, C., Gerard, C., Perret, J., Grootegoed, A. and Vassart, G. (1992) Expression of members of the putative olfactory receptor gene family in mammalian germ cells. *Nature* **355**, 453–455.
37. Thomas, M.B., Haines, S.L. and Akeson, R.A. (1996) Chemoreceptors expressed in taste, olfactory and male reproductive tissues. *Gene* **178**, 1–5.
38. Ning, Y., Roschke, A., Smith, A.C.M., Macha, M., Precht, K., Riethman, H., Ledbetter, D.H. (Group 1), and Flint, J., Horsley, S., Regan, R., Kearney, L., Knight, S., Kvaloy, K., Brown, W.R.A. (Group 2) (1996) A complete set of human telomeric probes and their clinical application. *Nature Genet.* **14**, 86–89.
39. Speed, R.M. (1988) The possible role of meiotic pairing anomalies in the atresia of human fetal oocytes. *Hum. Genet.* **78**, 260–266.
40. Flint, J., Wilkie, A.O., Buckle, V.J., Winter, R.M., Holland, A.J. and McDermid, H.E. (1995) The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation. *Nature Genet.* **9**, 132–140.
41. Vocero-Akbani, A., Halms, C., Wang, J.-C., Sanjurjo, F.J., Korte, Sarfate, J. *et al*. (1996) Mapping human telomere regions with YAC and P1 clones, chromosome-specific markers for 27 telomeres including 149 STSs and 24 polymorphisms for 14 proterminal regions. *Genomics* **36**, 492–506.
42. Rosenberg, M., Hui, L., Ma, J., Nusbaum, H.C., Clark, K., Robinson, L., Dziadzio, L., Swain, P.M., Keith, T., Hudson, T.J., Biesecker, L.G. and Flint, J. (1997) Characterization of short tandem repeats from thirty-one human telomeres. *Genome Res.* **7**, 917–923.
43. Saitou, N. and Ueda, S. (1994) Evolutionary rates of insertion and deletion in noncoding nucleotide sequences of primates. *Mol. Biol. Evol.* **11**, 504–512.
44. Efstradiadis, A., Posakony, J., Maniatis, T., Lawn, R., O'Connell, C., Spritz, R., DeRiel, J., Forget, B., Weissman, S., Slightom, J., Blechl, A., Smithies, O., Bralle, F., Shoulders, C. and Proudfoot, N. (1980) The structure and evolution of the human beta-globin gene family. *Cell* **21**, 653–668.
45. Eichler, E.E., Lu, F., Shen, Y., Antonacci, R., Jurecic, V., Doggett, N.A., Moyzis, R.E., Baldini, A., Gibbs, R.A. and Nelson, D.L. (1996) Duplication of a gene-rich cluster between 16p11.1 and Xq28: a novel pericentromeric-directed mechanism for paralogous genome evolution. *Hum. Mol. Genet.* **5**, 899–912.
46. Collura, R.V. and Stewart, C.B. (1995) Insertions and duplications of mtDNA in the nuclear genomes of Old world monkeys and hominoids. *Nature* **378**, 475–489.
47. van Deutekom, J.C.T., Bakker, E., Lemmers, R.J.L.F., van der Wielen, M.J.R., Bik, E., Hofker, M.H., Padberg, G.W. and Frants, R.R. (1996) Evidence for subtelomeric exchange of 3.3 kb tandemly repeated units between chromosomes 4q35 and 10q26: implications for genetic counselling and etiology of FSHD1. *Hum. Mol. Genet.* **5**, 1997–2003.
48. Lancet, D., Ben-Arie, N., Cohen, S., Gat, U., Gross-Isseroff, R., Horn-Saban, S., Khen, M., Lehrach, H., Natochin, M., North, M., Seidemann, E. and Walker, N. (1993) Olfactory receptors, transduction, diversity, human psychophysics, and genome analysis. In *The Molecular Basis of Smell and Taste Transduction*. Wiley, Chichester, Ciba Found. Symp. 179, pp. 131–146.
49. Rouquier, S., Taviaux, S., Trask, B.J., Brand-Arpon, V., van den Engh, G., Demaille, J. and Giorgi, D. Distribution of olfactory receptor genes in the human genome (submitted).
50. Upcroft, P., Chen, N. and Upcroft, J.A. (1997) Telomeric organization of a variable and inducible toxin gene family in the ancient eukaryote *Giardia duodenalis*. *Genome Res.* **7**, 37–46.
51. van der Ploeg, L.H., Gottesdiener, K. and Lee, M.G. (1992) Antigenetic variation in African trypanosomes. *Trends Genet.* **8**, 452–457.
52. Corcoran, L.M., Thompson, J.K., Walliker, D. and Kemp, D.J. (1988) Homologous recombination within subtelomeric repeat sequences generates chromosome size polymorphisms in *P.falciparum*. *Cell* **53**, 807–813.
53. Sullivan, S.L., Adamson, M.C., Ressler, K.J., Kozak, C.A. and Buck, L.B. (1996) The chromosomal distribution of mouse odorant receptor genes. *Proc. Natl Acad. Sci. USA* **93**, 84–888.
54. Elliott, R.W. and Pazik, J. (1995) An interstitial telomere array near Hba on mouse Chr 11 is a candidate for the homolog of the telomere at human 16p. *Genomics* **27**, 217–218.
55. Amann, J., Valentine, M., Kidd, V.J. and Lahti, J.M. (1996) Localization of Chl1-related helicase genes to human chromosome regions of 12p11 and 12p13, similarity between parts of these genes and conserved human telomeric-associated DNA. *Genomics* **32**, 260–265.
56. Riethman, H.C., Moyzis, R.K., Meyne, J., Burke, D.T. and Olson, M.V. (1989) Cloning human telomeric DNA fragments into *Saccharomyces cerevisiae* using a yeast-artificial-chromosome vector. *Proc. Natl Acad. Sci. USA* **86**, 6240–6244.
57. Kim U-J., Birren, B.W., Slepak, T., Mancino, V., Boysen, C., Kang, H.L., Simon, M.I. and Shizuya, H. (1996) Construction and characterization of a human bacterial artificial chromosomal library. *Genomics* **34**, 213–218.
58. Laboratory Manual of the Cytogenetics Division of the San Diego Zoo, Center for Reproduction of Endangered Species, Protocol 3.
59. de Jong, P.J., Yokabata, K., Chen, C., Lohman, F., Pederson, L., McNinch, J. and Van Dilla, M. (1989) Human chromosome-specific partial digest libraries in lambda and cosmid vectors. *Cytogenet. Cell Genet.* **51**, 585.
60. Trask, B.J., van den Engh, G.J., Christensen, M., Massa, H., Gray, J.W. and Van Dilla, M.V. (1991) Characterization of somatic cell hybrids by bivariate flow karyotyping and fluorescence *in situ* hybridization. *Somat. Cell Mol. Genet.* **17**, 117–136.
61. Trask, B.J. (1997) Fluorescence in situ hybridization. In Birren, B., Green, E., Hieter, P. and Myers, R. (eds), *Genome Analysis, A Laboratory Manual*. Cold Spring Harbor Laboratory Press (in press).
62. Telenius, H., Carter, N.P., Bebb, C.E., Nordenskjold, M., Ponder, B.A.J. and Tunnacliffe, A. (1992) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* **13**, 718–725.
63. http://ftp.genome.washington.edu/RM/RepeatMasker.html
64. http://serac.mbt.washington.edu/~chrisa/software/sputnik.html
65. http://www.genome.washington.edu/phrap_documentation.html
66. Wong, G.K.-S., Yu, J., Thayer, E.C. and Olson, M.V. (1997) Multiple-complete-digest restriction-fragment mapping: generating sequence-ready maps for large-scale DNA sequencing. *Proc. Natl Acad. Sci. USA* **94**, 5225–5230.
67. Sambrook, J., Fritsch, E.F. and Maniatis, R. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
68. Mefford, H., van den Engh, G., Friedman, C. and Trask, B.J. (1997) Analysis of the variation in chromosome size among diverse human populations by bivariate flow karyotyping. *Hum. Genet.* **100**, 138–144.
69. Goold, R.D., diSibio, G.L., Xu, H., Lang, D.B., Dadgar, J., Magrane, G.G., Dugaiczyk, A., Smith, K.A., Cox, D.R., Masters, S.B. *et al*. (1993) The development of sequence-tagged sites for human chromosome 4. *Hum. Mol. Genet.* **2**, 1271–1288.
70. Collard, J.G., de Boer. P.A., Janssen, J.W., Schijven, J.F. and de Jong, B. (1985) Gene mapping by chromosome spot hybridization. *Cytometry* **6**, 179–185.