 Human Molecular Genetics, 1998, Vol. 7, No. 13 **2007–2020**

ARTICLE

Large multi-chromosomal duplications encompass many members of the olfactory receptor gene family in the human genome

Barbara J. Trask*, Hillary Massa, Veronique Brand-Arpon[1], Kin Chan[2], Cynthia Friedman, Oanh T. Nguyen, Evan Eichler[3], Ger van den Engh, Sylvie Rouquier[1], Hiroaki Shizuya[2] and Dominique Giorgi[1]

Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195, USA, [1]Institut de Genetique Humaine, CNRS, Montpellier, France, [2]Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA and [3]Department of Genetics, Case Western Reserve University and University Hospitals of Cleveland, Cleveland, OH 44016, USA

Received July 29, 1998; Revised and Accepted September 28, 1998

The human genome contains thousands of genes that encode a diverse repertoire of odorant receptors (ORs). We report here on the identification and chromosomal localization of 74 OR-containing genomic clones. Using fluorescence *in situ* hybridization (FISH), we demonstrate a striking homology among a set of ~20 OR locations, illustrating a history of duplications that have distributed OR sequences across the genome. Half of the OR-containing BACs cloned from total genomic DNA and 86% of cosmids derived from chromosome 3 cross-hybridize to a subset of these locations, many to 17 of them. These paralogous regions are distributed on 13 chromosomes, and eight lie in terminal bands. By analyzing clones from an ~250 kb clone-walk across one of these sites (3p13), we show that the homology among these sites is extensive (>150 kb) and encompasses both OR genes and intergenic genomic sequences. The FISH signals appear significantly larger at some sites than at the native location, indicating that portions of some duplicons have undergone local amplification/attrition. More restricted duplications involving pairs of other genomic locations are detected with 12% of the OR-BACs. Only a small subset of OR locations is sufficiently diverged from the others that clones derived from them behave as single-copy FISH probes. We estimate that duplications encompassing members of the OR gene family account for >0.1% of the human genome. A comparison of FISH signals at orthologous locations in other primates indicates that a portion of this OR 'subgenome' has been in flux during the divergence of primates, possibly as a mechanism for evolving the repertoire of olfactory receptors.

INTRODUCTION

Humans can discriminate thousands of odors (1). This capability is due to the expression of a diverse repertoire of odorant receptors (OR) in the specialized sensory neurons in the olfactory neuroepithelium. The OR receptors are encoded in the genome by a large family of genes, whose coding regions are short (~1 kb) (2,3). ORs are members of the much larger family of G-protein-coupled receptors with seven transmembrane segments. Genes that are very closely related to ORs are expressed in the tongue and, surprisingly, in testes (4,5). A large subset of OR genes, ranging in similarity from 45 to 100%, can be amplified from the human genome using degenerate PCR primers in two highly conserved transmembrane regions (6). Remarkably, over two-thirds of the human OR sequences captured in this way are apparent pseudogenes (6).

Members of the OR gene family are distributed among many locations in the human genome (5–8). When a pool of OR-specific sequences was used as a probe for fluorescence *in situ* hybridization (FISH), signals were observed at >25 locations situated on all but a few chromosomes (6). PCR analyses of flow-sorted chromosomes confirmed the dispersed nature of the human OR family (6). Each OR location is likely to comprise multiple genes and/or pseudogenes (5,8,9; V. Brand-Arpon, S. Rouquier, H. Massa, P. de Jong, C. Ferraz, P.A. Ioannou, J.G. Demaille, B.J. Trask and D. Giorgi, submitted for publication).

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

Only a single OR is expressed in each neuron (9–13). Therefore, the multi-chromosomal distribution of OR genes presents a conundrum for understanding the transcriptional regulation of the OR gene family. The distributed nature of the human OR gene family is in sharp contrast to the organization of antigen-receptor gene families, where diversity is generated by joining elements from a single genomic cluster of possible components. For the OR family, transcriptional-control mechanisms must contend with the multiplicity of gene locations to ensure that only one OR is expressed in each neuron (and from a single allele) (10–13) and that each receptor is expressed in the appropriate zone within the neuroepithelium (9,11,12,14), while insuring that a diverse repertoire of receptors is expressed in the tissue as a whole. It is anticipated that a study of the genomic sequences surrounding transcribed OR genes will reveal how this set of demands is fulfilled at the molecular level.

The history of events leading to the multitude of OR-containing locations in the human genome and the plethora of pseudogenes is also insufficiently understood. The FISH results with OR-specific sequences (6) lead to two hypotheses about the evolutionary relationships among the many OR-containing locations. One possibility is that processed pseudogenes have been inserted into some sites by retrotransposition. In this case, homology would be limited to the transcribed portions of OR sequences. Alternatively, sites may be related as a consequence of interchromosomal duplications of large genomic segments. In this case, the homology is expected to extend to sequences that flank and/or lie between the OR genes. Given the many OR pseudogenes in the human genome, it is relevant to ask whether blocks of pseudogenes have been duplicated. Although several clusters of mouse OR genes are associated with large genomic duplications (15), and a subset of OR genes is repeated near multiple human telomeres within a much larger unit of duplication (7), the evolutionary relationships among most of the OR-containing sites in the genome are not yet known. The observation of extreme polymorphism among humans in the distribution of a subtelomeric block of OR genes (7) also raised the possibility that phenotypic variation could arise through variation in gene copy number or genomic context. It is therefore important to assess the plasticity and variability of the portion of the genome devoted to the OR family.

As a step towards understanding the evolution and transcriptional control of this gene family, we describe here the identification and cytogenetic characterization of genomic clones encompassing OR genes and pseudogenes. We show that the majority of OR clusters contains sequences that are shared by >15 sites in the genome. Only a few OR-containing regions, e.g. at 1q21–22, 1q44, 17p13 and 19p13.2, are sufficiently diverged from the others that clones derived from them hybridize to unique locations in the genome. We show that duplications of large genomic regions have accompanied the spread of OR genes to many locations. Finally, we demonstrate that the chromosomal distribution of OR duplicons has changed during the divergence of the great apes.

RESULTS

Many OR-containing clones hybridize to multiple genomic locations

Genomic clones containing OR-like sequences were identified from two clone libraries, a BAC library of ~80–100-kb inserts cloned from total human genomic DNA and a cosmid library constructed using DNA from flow-sorted chromosome 3. The initial screen was performed by hybridization with a complex pool of OR-specific sequences. This OR pool was PCR amplified from total human genomic DNA using degenerate primers recognizing evolutionarily conserved regions of the OR proteins (see Materials and Methods). Positives were confirmed by (i) PCR amplification of a product of the expected size using the degenerate OR primers, (ii) sequencing of cloned PCR products and/or (iii) Southern blot hybridization of *Eco*RI-digested DNA using the OR sequence pool as a probe (data not shown). Of the clearly positive clones, 52 BACs and 22 chromosome 3 cosmids were characterized cytogenetically by FISH.

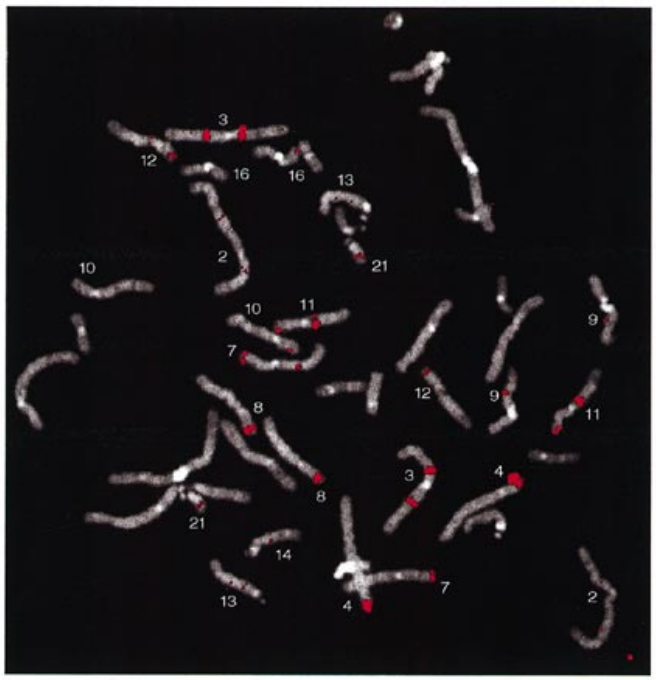

Figure 1. A metaphase spread showing the hybridization signals produced by the OR-containing BAC 51D11. By analyzing 10 such metaphases, it was possible to discriminate recurrent sites of specific hybridization from sporadic background signals. Twenty sites of specific hybridization were detected (Fig. 2A) on the 13 indicated chromosomes. More than one region of hybridization is evident on several chromosomes (2, 3, 7, 11, 12 and 13). Note the marked difference in signal intensity among the locations, which is reflected in the average signal scores that are plotted in Figure 2A.

Over half of the 74 clones produced FISH signals at more than one genomic location. Figure 1 illustrates the most striking multi-chromosomal pattern we observed: BAC 51D11 cross-hybridizes to 20 locations on a total of 13 chromosomes. The signal intensity and labeling efficiency varies significantly among the sites. The FISH signals at some sites, such as 4p16 and 8p23, exceed the intensity expected for a BAC of this size, suggesting that all or part of the BAC's insert has been duplicated locally at these locations. In contrast, signals at other sites, such as 2p12–13, 2q22–23, 13q13, 13q21 and 14q21, are relatively dim and are not seen in all metaphases, suggesting that these sites harbor sequences that are less homologous to or homologous to only a portion of the BAC's sequence.

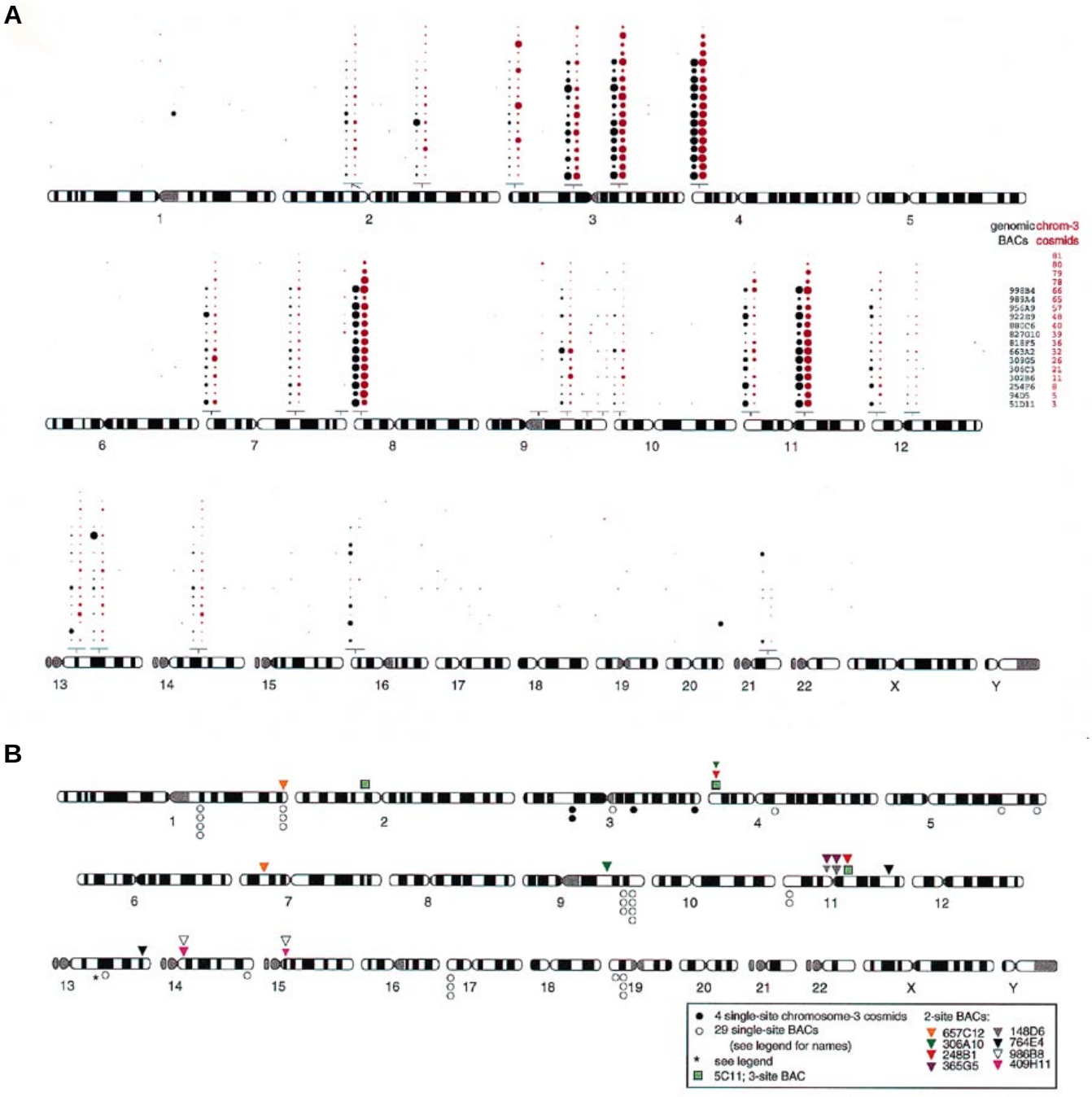

Figure 2. Summary of the hybridization signals produced by 74 OR-containing clones identified in a total genomic BAC library and a chromosome 3 specific cosmid library. (**A**) 32 clones produced signals consistently at 17 sites in the genome, and their signals are summarized in this panel. The black dots summarize the results of 14 OR-BACs, and the red dots summarize the results of 18 OR-containing chromosome 3 cosmids. Each row of dots corresponds to the signals observed by a particular clone. The size of each dot indicates the average signal intensity observed at a given location with a particular clone. The names of the clones are indicated in the two columns to the right. A small T is drawn at the 24 locations where signals were observed with more than four of the clones. Polymorphism (see text) may partially account for ambiguity in the assignment of clones in 2p. The stacks are positioned over the center of the hybridizing regions; the mapping precision is approximately half of a 400-level band. (**B**) Summary of the hybridization sites of BACs or cosmids that hybridized to three or fewer locations in the genome. Twenty-nine BACs hybridized to single sites; these sites are marked with white circles. Each circle denotes a separate clone, i.e. four different clones mapped to 1q21–22. Four chromosome 3 specific cosmids mapped to single sites, which are marked with black circles. Eight clones mapped to two sites, indicated with pairs of matching triangles. One clone mapped to the three sites indicated with a green square. The relative sizes of the triangles and square reflect the relative hybridization efficiency/intensity at the two or three sites, respectively. The names of the clones mapping to multiple locations are indicated in the figure. The single-locus BACs, listed in order of location, are: 1q21–22: 267C3, 460D10, 821D9, 980D1; 1q43–44: 176F8, 850H7, 992G12; 3q11.2–13.1: 748C6; 4q13–21: 984H6; 5q23–31: 995D3; 5q34: 303F10; 9q32-proximal q34: 17E12, 378E10, 855A10, 963F3, 987D11, 966G7, 996G8; 11p15: 626C11, 978C7; 13q21: 969B7 [*, 13q21 is the predominant site of 969B7, but this clone also produced dim signals infrequently at some of the same locations as the multi-site clones in (A)]; 14q32: 858F6; 17p12–13: 45F12, 284E5, 588A4; 19p13.3: 32E9; 19p13.1–13.2: 3F6, 272A4, 378D8. The single-locus chromosome 3 specific cosmids are 3p14–21: 13, 54; 3q13.3–21: 28; 3q28–29: 45. (The LL03NC01 designations are given in Materials and Methods.)

The multi-chromosomal distribution of BAC 51D11 is typical of 14 (26%) BACs and 18 (82%) chromosome 3 cosmids. Figure 2A summarizes the relative intensities of signals observed with these 32 multi-site clones. The relative average intensities of the signals at each location, indicated by the size of the dots representing each clone, generally reflect the pattern shown for 51D11. A total of 18 locations cross-hybridized to each of 10 cosmids and eight BACs (>50% of the multi-site clones of each type). This list includes locations on chromosomes 2p, 2q, 3p (two sites), 3q, 4p, 7p, 7q, 8p, 9q, 10p, 11p, 11q, 12p, 12q, 13q (two sites) and 14q. In addition, more than eight BACs produced a strong FISH signal at 16p, and several cross-hybridized intensely to 21q (51D11 in Fig. 1 labels both these locations). Twenty-four sites showed hybridization signals with at least four of the clones, and eight sites were detected with all 32 clones. Of the 24 sites of significant homology, 42% are located in terminal bands, which together account for only ~10% of the genome.

The remaining 38 OR-BACs and four OR-containing chromosome 3 cosmids exhibited more restricted chromosomal distributions than the set typified by 51D11. The FISH results of these clones are summarized in Figure 2B. Many clones (12 BACs and one cosmid) hybridized to a subset of the locations (defined as within the resolution of two-color metaphase FISH) observed with the multi(>15)-site clones. 5C11 hybridized to three of these sites. Interestingly, 5C11 labels 2p12–13 as brightly as it does 4p16 and 11q12–13, whereas 2p12–13 was usually one of the dimmer locations seen with the set of clones that cross-hybridize to many more locations than 5C11. Similarly, 969B7 produced a predominant signal at 13q21, but dim signals were seen rarely at many of the same locations seen with the multi-site clones. We conclude that 5C11 and 969B7 derive from 2p12–13 and 13q21, respectively, and each contains a portion of the larger, widely distributed duplication.

Eight BAC clones produced signals at two locations. Two patterns are worth drawing attention to because each was observed with two BACs. An OR-related segment, represented by clones 365G5 and 148D6 (grey and dark purple triangles in Fig. 2B), appears to be duplicated on either side of the chromosome-11 centromere at locations distinct from the sites detected with the multi-site clones. Another pericentromeric duplication involving 14q11.2 and 15q11.2–12 was detected with OR-containing clones 986B9 and 409H11 (white and pink triangles). In addition, the pairs of sites labeled with 248B1 and 306A10 (red and green triangles) are subsets of the sites detected by the multi-site clones. The results for the remaining two 2-site clones are shown in Figure 2B (orange and black triangles) for completeness, but the possibility has not been excluded that they represent cloning artifacts rather than bona fide genomic duplications.

Finally, four (18%) of the OR chromosome 3 cosmids and 28 (54%) of the OR-BACs behaved as simple single-copy probes (black and white circles, respectively, in Fig. 2B). Seven locations, 1q21–22, 1q43–44, 3p14–21, 9q32–34, 11p15, 17p13 and 19p13.2, were identified with two or more clones. OR gene clusters at the latter three locations are already the subjects of detailed molecular analyses (see Discussion). Approximately 10 of the 16 locations identified with these single-copy clones were identified previously with OR-specific probes (6). Of these 16 locations, five were identified by one or more of the 2-, 3- or multi-site clones, suggesting again that some OR-containing regions are a composite of relatively unique segments and duplicated segments.

Characterization of the extent and structure of the OR-containing duplications

We wished to determine if the duplications that resulted in homology among the 17+ sites, detected by the clones shown in Figures 1 and 2A, involved sequences outside of the OR genes themselves. We had shown previously that OR sequences reside at most of these locations (6), indicating that the OR sequences are part of the duplications. We conclude from the following six observations that non-OR sequences are also part of the paralogous segments.

(i) The intensity of the FISH signals at most of the cross-hybridizing locations was significantly greater with the genomic clones than with short OR-specific probes.

(ii) The addition of excess unlabeled OR-sequences to biotinylated cosmid 8 failed to attenuate the resulting FISH signals significantly at any of the locations (data not shown).

(iii) PCR assays designed from the sequences at the ends of several cosmids confirm the multiplicity of their cytogenetic locations. These primers were designed to avoid common interspersed repeats and OR homology. Each of the 24 human chromosomes, isolated in monochromosomal rodent somatic cell hybrid lines, was subjected to PCR amplification with these primer-pairs. In contrast to FISH, PCR assays chromosomes for a small region of a clone and demands that the sequence be duplicated with sufficient similarity that PCR-amplification occurs. As expected, each PCR assay yielded a product on chromosome 3, the origin of the cosmids. Primers designed from the T7 end of clone 32 [which produced a FISH signal on chromosomes 2–4, 7–14 and 21 (Fig. 2A)] amplified a product of the expected size from hybrids containing human chromosomes 2–4, 7–11 and 13 (Fig. 3A). Only chromosomes 12, 14 and 21, which were seen by FISH, failed to amplify. Chromosomes 3, 4 and 8 were positive for a PCR assay designed at the opposite, T3 end of this clone (not shown). Primers from the T7 end of cosmid 5 amplified a product from chromosomes 2–4 and 13 in the panel, and those from the other end amplified from chromosomes 3 and 6 (not shown). Primers matching the T3 end of cosmid 11 amplified a strong product from chromosomes 3, 4, 7–13 and X (not shown). FISH of this clone detected homologous sequence on all these chromosomes except X, plus significant signals on chromosomes 2 and 14 (Fig. 2A). Thus, these PCR assays confirm the FISH findings and demonstrate that the homologous segments extend outside of the OR genes and can encompass sequences at both ends of a cosmid insert (as in the case of cosmid 32 on chromosomes 3, 4 and 8).

(iv) FISH analyses of clones derived from a 250 kb contig from 3p13 encompassing several OR genes (V. Brand-Arpon *et al*., submitted for publication) allowed us to characterize the extent and coarse structure of the homologous duplications. The FISH results of eight clones spanning this contig are summarized in Figure 4. Only some of these clones contain OR sequences (Fig. 5A). These are PAC 169, cosmids 26, 3, 48 and 81. Clone 88 overlaps PAC 169, but contains no OR genes. Clones 96 and 97 contain no OR sequences, but encompass a MLCK pseudogene (V. Brand-Arpon *et al*., submitted for publication). Sequences sampled from the middle of the contig, i.e. in cosmids 26, 3 and

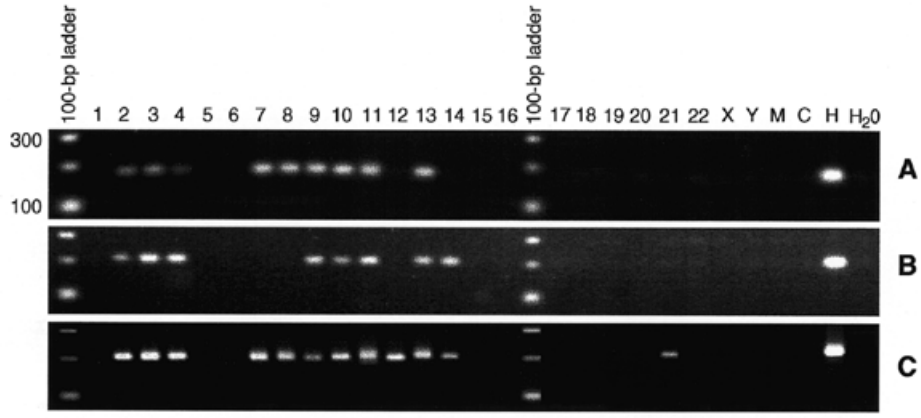

Figure 3. PCR analyses of non-OR portions of clones that FISH-map to multiple locations. (**A**) Primer pair 32MF/32IR designed from the T7 end of cosmid 32, whose FISH distribution is summarized in Figure 2A. (**B**) Primer pair U45021 and L45210, designed from sequence 5′ of the cluster of OR sequences in the 3p13 contig (PCR assay D in Figure 5). (**C**) Primer pair U66952 and L67132, designed from the sequence lying between two OR sequences in the 3p13 contig (PCR assay E in Figure 5). The FISH results of cosmids 3, 48 and 81, which overlap one or both of the latter two assays, are summarized in Figures 2A and 4. The lanes are numbered to indicate the chromosomes contained in each hybrid cell line. M, mouse genomic DNA; C, Chinese hamster genomic DNA; H, human genomic DNA; H_2O, water control.

48, are shared by ~17 sites. Sequences in cosmids 96 and 97 are duplicated at only two of these locations, 3p13 and 3q13–21. Clone 81, which overlaps both of these two groups of clones, hybridizes predominantly to 3p13 and 3q13–21, but dim signals are also observed infrequently at several other locations. Clones at the left end of the contig hybridize to an extensive set of chromosomal locations, which overlaps but is slightly different than the set observed with clones at the middle of the contig. Signals become dimmer and less frequent on chromosome 2 (both locations), 3p, 13q and 14q, but signals are relatively brighter on 12p, and homology is additionally detected on 16p, 19p, 19q and 21q. This pattern is borne out by FISH analyses of eight additional clones that were isolated by screening the chromosome 3 library with the distal ends of PAC 169 (cosmids 202, 204, 209, 215 and 219) and cosmid 88 (cosmids 111, 113 and 117) (Fig. 5A). The origin of these clones has not been determined; they may derive from any of the three cross-hybridizing locations on chromosome 3. They lack OR genes (V. Brand-Arpon, unpublished data). The FISH results with these clones demonstrate that the paralogy continues beyond the end of PAC 169. Clones 111, 113 and 117 produced signals on chromosome 3 (three sites), 4p, 7p, 7q, 8p, 9q, 10p, 11p, 11q, 12p, 12q and 16p, but only rarely on chromosomes 2, 13 and 14. Clones 202–219 produced the same pattern as PAC 169, but additionally labeled a fourth site on chromosome 3, at 3q28–29. Thus, a compilation of the physical map of the 3p13-contig and the FISH results demonstrates that (a) the region can be coarsely subdivided into four zones with different hybridization patterns involving overlapping sets of locations, (b) the homology between two sites on chromosome 3 includes sequences spanning >250 kb, and (c) sequences shared by >10 chromosomes span at least 150 kb.

(v) PCR assays confirm the multi-chromosomal distribution of sequences across the 3p13 contig (Fig. 5B). Most notable are PCR assays D and E, which lie just outside of and between a cluster of OR genes, respectively (Fig. 3B and C). Products of the predicted size are generated for assay D on most, and for E on all of the chromosomes where FISH signals were observed using overlapping cosmids (Fig. 3B and C). Primers at the T7 end of cosmid 88, which lie ~50 kb from this OR cluster and an undetermined distance from the OR gene in PAC 169, amplify a product of the same size from chromosomes 3, 4, 10, 11 and 12 (Fig. 5B). These sites are a subset of the sites detected by FISH using overlapping clones (Fig. 4). PCR assays at the T3 end of cosmid 88 amplify only from chromosomes 3 and 9. As expected, PCR assays at the MLCK end of the contig in cosmids 96 and 97 amplify from chromosome 3 only.

(vi) Sequence comparisons demonstrate that ~30 kb of the 3p13 sequence (DDBJ/EMBL/GenBank accession no. AF042089; V. Brand-Arpon *et al*., submitted for publication) is duplicated within a PAC clone (PAC pDJ392a17) that has been assigned to chromosome 11 (G.A. Evans *et al*., unpublished; GenBank accession no. AC000385) (Fig. 6). The PAC contains three OR sequences, which are similar but not identical to the three OR sequences in the 3p13 sequence. The dot-matrix comparison in Figure 6 shows that the homology extends beyond these OR sequences. A larger duplication unit of ~20–24 kb can be discerned within the homologous region. This unit encompasses two clusters of many copies of an imperfect ~63 bp repeat (termed VNTR for convenience), a retroviral pGAG element, and two OR genes separated by a region (denoted in turquoise) composed of 'unique' sequence and interspersed repeats. The PAC contains almost two complete copies of this duplication unit encompassing 43.9 kb; the 3p13 sequence contains approximately one and one-third copies spanning ~30 kb. In both cases, the units are arranged head-to-tail. The VNTR blocks vary in size (2.3–5 kb) and structure among the paralogous copies in these two clones, and Alu and L1 insertions distinguish the regions of homology between the OR genes. The longer regions of high homology are ~90% identical. Corresponding OR pseudogenes in the two regions (i.e. at the same position relative to the VNTR-blocks and GAG elements in each duplication unit) share many of the same deleterious mutations. Of the 20 positions where frame-shifts, in-frame stops or faulty start/termination signals are identified among these six ORs, nine are seen in more than one OR (five at the same relative position in both PAC and cosmids). Homology ends just 5′ of the position of the pMLCK sequence in the 3p13

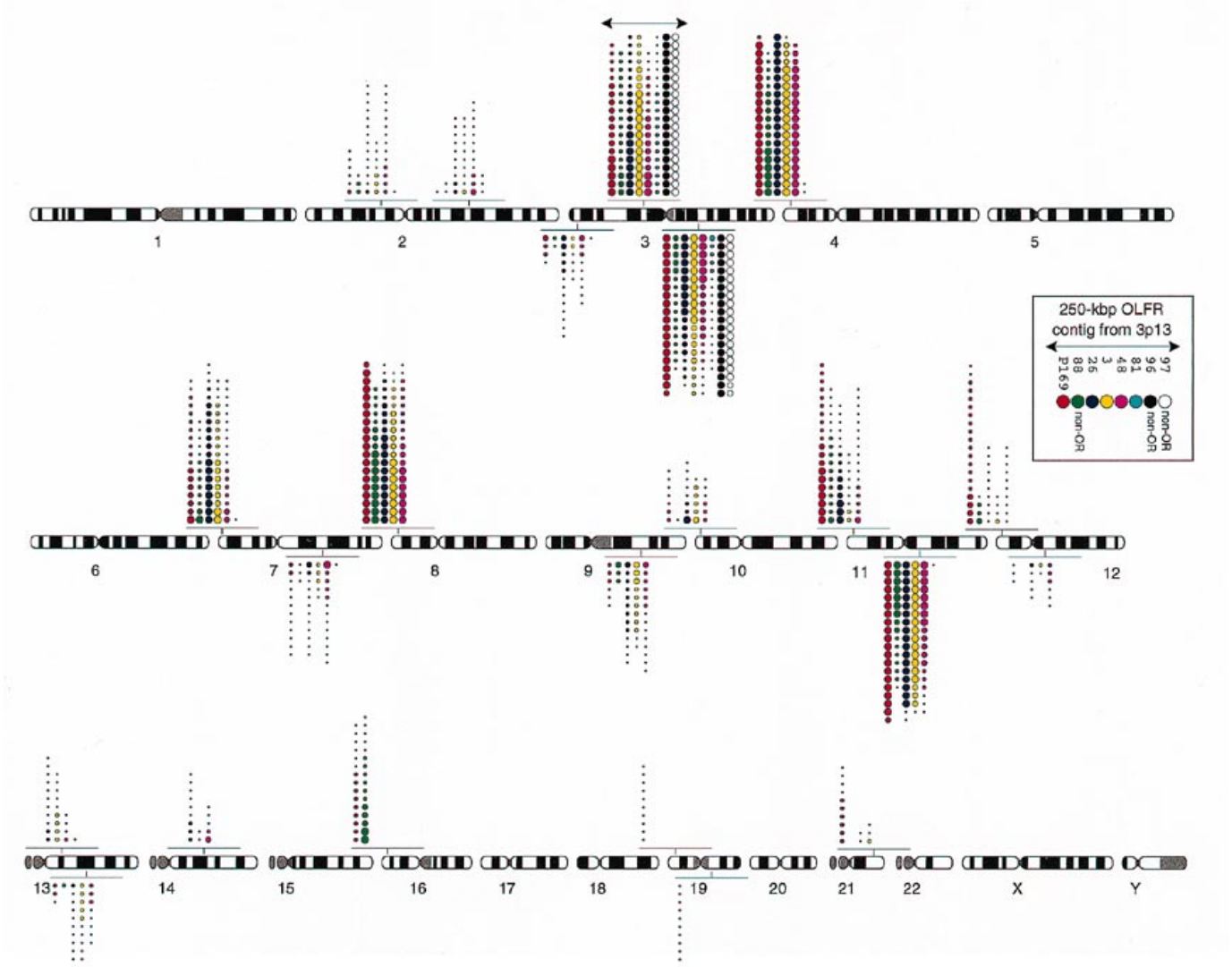

Figure 4. FISH results of eight clones derived from 3p13 and distributed across a 250 kb contig (V. Brand-Arpon *et al*., submitted for publication). The position of each clone is indicated in Figure 5A. The FISH results of each clone are summarized with a different color. Each stack of dots indicates the relative signal intensity observed in each of 20 chromosomes in 10 metaphases. The diameter of the dots is proportional to the signal intensity, which was scored on a scale of 0–4. A stack of 20 dots signifies that signals were observed on all analyzed chromosomes at that site; shorter stacks indicate hybridization efficiencies <100%. Each group of stacks is aligned with the midpoint of the hybridization range. Signals could typically be assigned with a precision of half of a 400-level band. Note that the black and white stacks for cosmids 96 and 97 are confined to two locations on chromosome 3, and PAC 169 and cosmid 88 (red and green stacks, respectively) show a different pattern than do cosmids 3, 26 and 48.

sequence (at position 72 660). Because the PAC's sequence ends at position 37 360 in the 3p13 sequence, the structure of the homology extending 5′ of the OR sequences cannot be characterized.

Large-scale heteromorphism of these OR clusters is minimal in humans

Because we had previously observed striking polymorphism in copy number and chromosomal location of one cluster of OR genes that map exclusively near telomeres (7), we analyzed the FISH patterns of 17 of the OR clones on two or more individuals. This set includes 10 OR-BACs that hybridized to a single location and five OR-BACs and three OR-cosmids that hybridized to multiple sites. With one exception, the patterns of hybridization were very similar on the different individuals. Slight differences could be ascribed to differences in hybridization efficiency and statistical sampling. We observed only a subtle polymorphism in the hybridization pattern on the p-arm of chromosome 2. Several of the multi-site clones produced signals at two locations within the p11.2–p13 interval on chromosomes of two individuals, while only one or the other site was labeled in other individuals. (For simplicity, signals observed at both locations are lumped into a single group in the clone summary in Fig. 2A.)

Differences among primates in the genomic organization of OR duplications

In order to judge the plasticity in the cytogenetic organization of OR-containing duplications, we compared the FISH patterns of

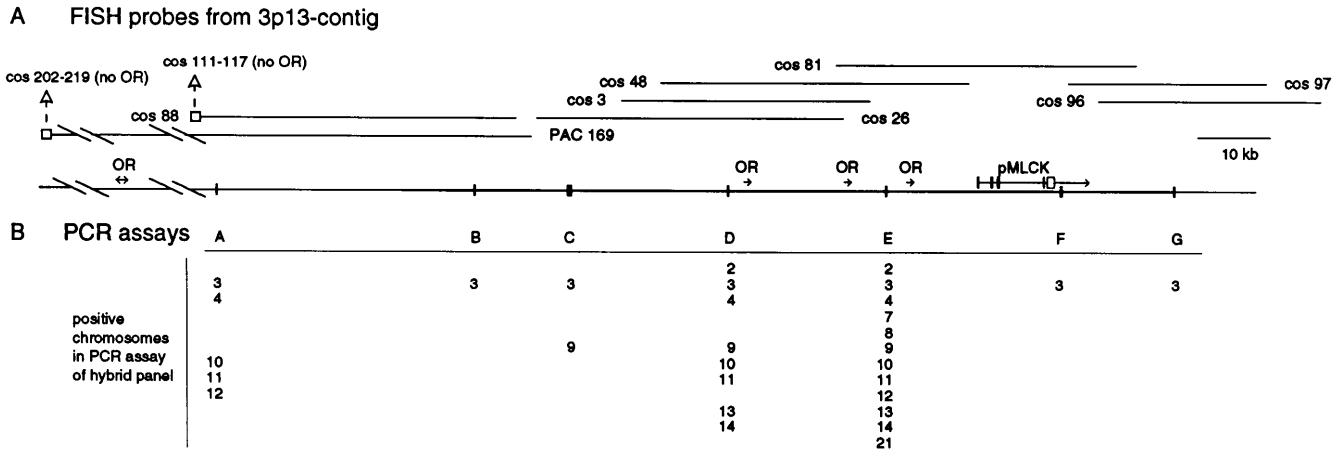

Figure 5. (**A**) A simplified map of a 3p13 contig encompassing four OR sequences (V. Brand-Arpon *et al*., submitted for publication). The locations of the eight clones spanning the contig that were analyzed by FISH are indicated by the horizontal lines. In addition, we also FISH-mapped three clones that were identified by screening the chromosome 3 cosmid library with the T7 end of cosmid 88 (cosmids 111, 113 and 117) and four clones identified by screening the library with the T7 end of PAC 169 (cosmids 202, 204, 209, 215 and 219). These clones may derive from paralogous regions elsewhere on chromosome 3, and their overlap has not yet been established. These clones and cosmids 88, 96 and 97 lack detectable OR sequences. The thick line designates the 106 kb region that has been sequenced (V. Brand-Arpon *et al*., submitted for publication). The locations of the four OR sequences and an MLCK pseudogene are indicated by the small arrows. The OR sequence in PAC 169 was identified by PCR amplification of the PAC with the OR3B/OR5B primers. Its orientation and exact position is not known, but it appears to lie in the region not overlapped by cosmids 202–219, 88 or 111–117, because the OR3B/5B primers do not amplify OR sequences from these clones. (The LL03NC01 designations for these clones are given in Materials and Methods.) (**B**) PCR results using PCR–primer pairs designed from the 106 kb sequenced contig of 3p13 (assays C–F) and from the end-sequences of clones in the contig (A, B and G) (see Materials and Methods for details). The positions of the primer pairs are indicated in (A). All chromosomes in the Coriell monochromosomal panel were assayed; only the positive chromosomes are listed in the table. Gels showing assays D and E are shown in Figure 3B and C, respectively.

two of the multi-site OR clones in human, chimpanzee, gorilla, orangutan, gibbon and baboon. The results with cosmid 3 in the first four species are summarized in Figure 7. Cosmid 3 lies in the middle of the 3p13-contig and contains two OR sequences (Fig. 5A). Despite the general conservation in banding patterns among these four species, a variety of gross changes have occurred with the sequences cross-hybridizing to this clone. Of the many differences evident in Figure 7, the following are considered significant because they were replicated with a different OR-containing clone (94D5) and/or cosmid 202, which was identified by walking 5′ of the OR genes in the 3p13 contig. For simplicity, the relevant locations are identified here by their orthologous location in the human karyotype. The most striking changes are alterations in the position of cross-hybridizing sequences on chromosome 2, the lack of signal at 3q13–21 in orangutan, the lack of cross-hybridizing sequences in the 4p region in orangutan and a diminution of the signal at this location in chimpanzee (but retention in gorilla), loss of the 7p signal in gorilla although it is present at this location in the other great-apes, lack of both sites on 11 in orangutan, and a bright region of cross-hybridization on 16p in orangutan, but not in the other great-apes.

We used cosmid 202 to analyze the duplications in gibbon (*Hylobates lar*). This clone produces signals at six locations, on gibbon chromosomes 1, 4, 7, 8, 11 and 12. These sites correspond to regions on human chromosomes 7, 3, 8, 16, 11 and 3, respectively (16; data not shown). In contrast, the brightest and/or most frequently labeled locations in human are two sites on 3, 4p16, 8p, 11p, 11q and 16p (plus dimmer and less frequent signals on 3, 7 and 12). Thus, like orangutan, gibbon lacks cross-hybridizing sequences at regions corresponding to HSA 4p16, one of the brightest locations in the human genome.

The sacred baboon (*Papio hamadryas*) is more distantly related to human than is gibbon, yet chromosome-painting experiments suggest that fewer chromosomal rearrangements exchanges have occurred during its divergence from human than have occurred along the branches separating gibbon and human (17). Cosmid 202 produced signals at at least eight locations on baboon chromosomes. The brightest lie in regions corresponding to HSA 1 (two sites), 3, 7, 16 and 19. No signals were observed in regions thought to correspond to HSA 4, 8 or 11 (PHA 5, 8 or 14), the most prominent locations in human.

DISCUSSION

Large duplications in the human genome

It has long been recognized that genomes are shaped by evolutionary processes that include the duplication of chromosomal segments (18). As a consequence of these past duplications, the human genome is riddled with repeats, ranging from the small and ubiquitous Alu-elements to large low-copy repeats (7,9–30 and references therein). The scope and complexity of the genomic duplications involving the OR family reported here are unprecedented, however, and are perhaps matched only by the extensive paralogy of regions near telomeres (30), where OR genes have also been found (7). Of the 44 locations identified here with OR-containing genomic clones, all but 11 (24%) contain sequences that cross-hybridize to one or more other locations in the genome. [If we consider only the 32 locations identified with two or more OR-containing clones, all but 5 (16%) contain sequences related to at least one other location in the genome.] Most of these regions were not previously known to have close

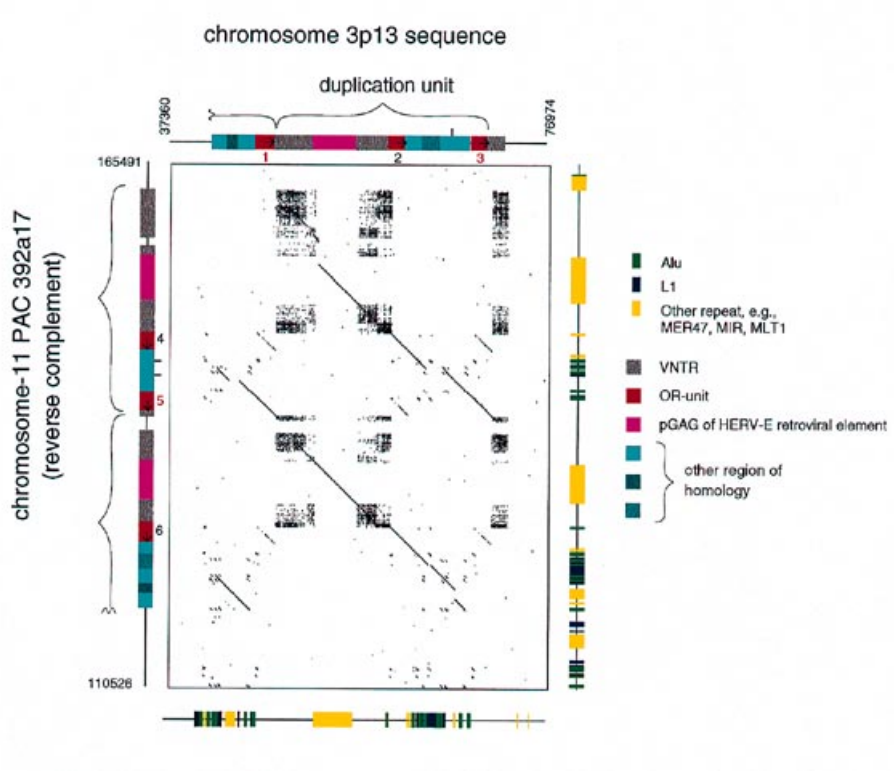

Figure 6. A dot-matrix comparison of homologous regions in 106 kb sequence from 3p13 (GenBank accession no. AF042089) and the reverse-complement of the sequence of PAC 392a17 (G.A. Evans *et al*., unpublished data; GenBank accession no. AC000385), which has been assigned to chromosome 11. The matrix was generated using the ABI Inherit program, using the following parameter settings: window size, 20; offset, 10; match, 85%. Diagonal lines in the matrix indicate regions of homology between the two sequences. A 20–24 kb duplication unit composed of several distinct homology units can be discerned and is denoted by the colored segments above and to the right of the matrix. The 12 conspicuous blocks of sequence matches correspond to alignments of the three VNTR-like regions of the 106 kb sequence (at approximate nucleotide positions 48330–52320, 56950–60315 and 70530–72620 in AF042089) with the four VNTR regions in the PAC (designated by the grey blocks). These VNTRs are composed of imperfect repeats of a unit averaging 63 bp in length (the largest block in the 3p13 sequence contains ~40 copies of this repeat). A portion of the GAG region of an HERV-E endogenous retroviral element comprises the homologous segment indicated in purple. The regions corresponding to the OR sequences are indicated in red. The ORs are numbered from 1 to 6, with the red and black colors of the numbers indicating the two groups of three highly homologous ORs. The regions denoted in turquoise are ~90% identical, but their structure varies due to a variety of Alu and L1 insertions (darker tones of turquoise). The positions of common interspersed repeats detected with RepeatMasker (http://ftp.genome.washington.edu/RM/RepeatMasker.html) are indicated along the bottom of the dot-matrix for the 3p13 sequence and along the right edge for the PAC. The duplication unit is not repeated in the regions of the 3p13 sequence 5′ and 3′ of the segment shown in the dot-matrix or elsewhere in the PAC. The PAC's sequence ends at position 165491.

relatives elsewhere in the genome. Of the 74 OR-containing BACs and cosmids we characterized, 57% hybridize to two or more locations. The similarity of these interchromosomal duplications is sufficiently high (~90%) and extensive (>150 kb) that cross-hybridization is detected under conventional FISH conditions. Note that hybridization of Alu and L1 elements and other high-copy repeats to the chromosomes is blocked in these experiments. Our most striking finding is the sequence homology among a set of >20 locations distributed on 13 chromosomes (Fig. 2A). We refer to the 32 genomic clones cross-hybridizing to at least 15 of these sites as 'multi-site clones'.

Homology among >15 locations comprises both OR and non-OR sequences

Several lines of evidence lead us to conclude that the sequences common to these many locations comprise a combination of OR and non-OR sequences. Pools of small OR-specific probes produced significant accumulations of FISH signals (6) at or near 21 of the 24 locations detected here with four or more multi-site clones (within the error of FISH localizations carried out in separate laboratories). Thus, OR sequences lie in most, if not all, of these related locations. However, the FISH signals with the multi-site clones were significantly more robust than those produced by the OR-specific probes, and this pattern was not significantly altered by suppression with unlabeled OR sequences. Clones spanning ~150 kb of a 250 kb contig of 3p13 (V. Brand-Arpon *et al*., submitted for publication) cross-hybridize to 17 locations. One of these clones, cosmid 88, lacks OR sequences, yet its multi-site pattern is similar to those of its OR-containing neighbors. PCR assays designed from the ends of various multi-site clones as well as in OR-free regions of the 3p13-contig demonstrate that many of these paralogous duplications include non-OR sequences that are sufficiently similar to serve as a template for PCR.

The molecular structures of paralogous regions

The structures of the paralogous regions are intriguingly complex. Clones derived from a cosmid-walk across an OR-cluster on 3p13 allow us to divide this region coarsely into four zones, each with a different chromosomal distribution. The region 3′ of the cluster of OR sequences is duplicated on two sites on chromosome 3. The region encompassing the three OR genes appears to

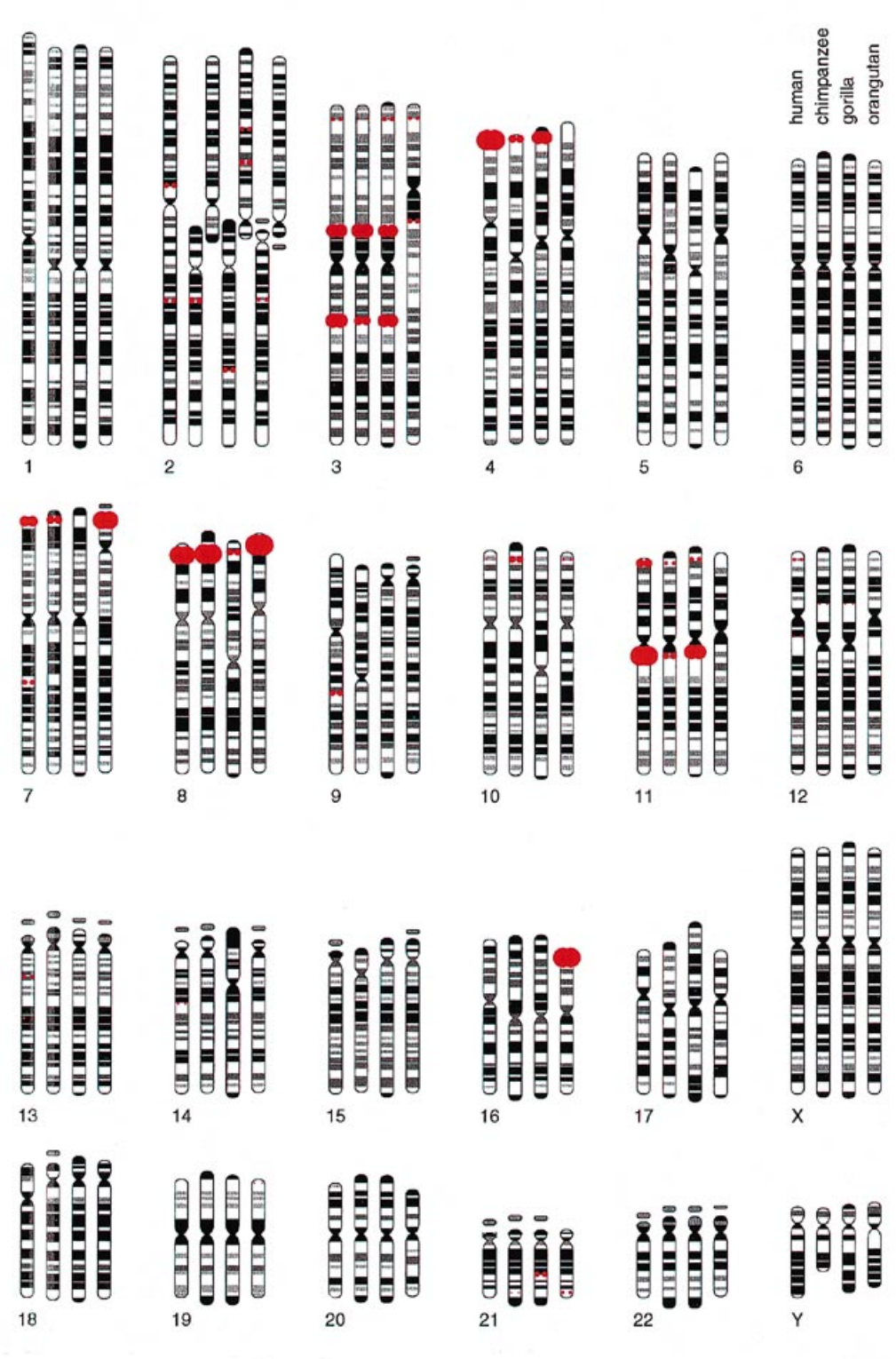

Figure 7. A comparison of the FISH results with OR-containing cosmid 3 (LL03NC01-4B17) on human, chimpanzee, gorilla and orangutan chromosomes. The chromosomes are numbered according to the human karyotype. Cosmid 3 is part of the 3p13 contig (Figure 5). For each species, 10 metaphase spreads were analyzed and signals at each location were scored on scale of 1–4. The size of the symbols is proportional to the average score in 20 chromosomes, and thus reflects combination signal intensity and hybridization efficiency. Differences in the FISH patterns that were observed with two or more tested clones are discussed in the text.

be duplicated at least in part on 18 sites. By walking 5′ of the OR cluster, a third zone is reached that is duplicated in several additional locations (16p, 19p, 19q and 21q), but is less homologous to some regions (2, 13 and 14). Clones identified by screening the chromosome 3 library with the 5′ most distal end of the 3p13 contig identify a fourth zone. Sequences in this zone hybridize to 3q28–29 in addition to all the sites identified by the third zone except those on 19 and 21. Although its structure and sequence variants distinguish the 3p13 region from its paralogs (V. Brand-Arpon *et al*., submitted for publication), our cosmid walks have so far failed to reach the boundaries of the duplicated region.

A comparison of the sequence of a PAC ascribed to chromosome 11 and 106 kb of 3p13 sequence allows us to describe the structure of the second zone of homology within the 3p13 contig in more detail (Fig. 6). Both regions contain three OR pseudogenes, whose coding portions account for only ~3 kb of the region they have in common. These sequences lie in a larger duplication unit of ~22 kb. A striking feature of these units is two 2–6 kb blocks of a characteristic VNTR-like sequence. Each repeat is a variation of an ~63 bp motif. A portion of a GAG element of a HERV-E endogenous retroviral element and two OR sequences separated by a non-descript 6–7 kb segment make up the remainder of the duplication unit. The PAC and the 3p13-contig contain one complete and 1–2 partial copies of this unit arranged head-to-tail. As a consequence of this arrangement, the first and third OR sequences in the 3p13 cluster and the second in the PAC are most closely related, and the second OR in the 3p13 cluster is most closely related to the first and third ORs in the PAC (see also below). We predict that homology between 3p13 and chromosome 11 will extend further 5′ of the region cloned in the PAC, because clones extending beyond this point in the 3p13 contig continue to cross-hybridize to multiple chromosomes, including two sites on chromosome 11.

We remarked that clones derived from 3p13 produced signals at many sites, such as 3q13–21, 4p16, 7p22, 8p23, 11p15, 11q12–13 and 16p13.3, that were as bright or brighter than the signals at the native 3p13 location. These results indicate that portions of the homologous region are likely to exist in multiple copies at these locations. In contrast, several sites, such as on chromosomes 2, 13 and 14, were labeled only dimly and infrequently with the multi-site clones. These sites also showed little to no cross-hybridization with clones lying 5′ of the cluster of 3 OR sequences in the 3p13 contig. These observations, combined with the positive results on these chromosomes for PCR assays that lie between the OR sequences within the 22 kb unit, suggest that homology at these sites is restricted to one or a few copies of this duplication unit.

Several regions appear to be a composite of OR-containing segments with many relatives in the genome and single-copy or more restricted OR-containing segments. For example, 11p15, 9q32–34 and 13q21 contain sequences that cross-hybridize to many multi-site clones, as well as segments that behave as single-copy probes (compare Fig. 2A and B). BAC 306A10 illustrates that a portion of the OR clusters on 9q22 and 4p16 is shared by only these two regions. Similarly, sequences in 248B1 are common to 11q12–13 and 4p16, but lie near a sequence shared by many additional locations.

Atypical intensities of several of the multi-site clones at particular locations belie their genomic origin. For example, clone 5C11 hybridizes particularly intensely to 2p12–13, which is dimly labeled by most of the other clones, 663A2 gives a very bright signal at 2q23, 989A4 at 13q21, and 922E9 at 21q22. It is likely that these clones contain sequences that overlap the duplicated region, but extend into relatively unique sequence at these locations.

Plasticity in the cytogenetic arrangement of OR duplications in primates

Previously, we detected marked polymorphism among humans using a subtelomeric block of OR genes, with copy-number ranging from 7 to 11 and wide diversity in chromosome locations (7). In contrast, we observe little large-scale polymorphism among humans using the subset of the OR family surveyed here.

However, our data show that much of the OR subgenome has been in flux at the cytological level during the divergence of primates. Only chimpanzee and human show a similar pattern of duplications with the multi-site clones. The most significant difference between these species is diminution of the signal on 4p in chimpanzee relative to human and gorilla.

Our results illustrate that many more molecular changes distinguish the genomes of primates than are evident from the conservation in banding patterns among the great apes (31) or the contiguity of regions labeled with chromosome-paints in baboon. A variety of changes in location and intensity of OR-containing segments have occurred during primate evolution. Of the six primates analyzed, only three (human, chimpanzee and gorilla, have sequences cross-hybridizing to the multi-site clones on 4p. Signal is lacking in these same species on 16p, where a large block of cross-hybridizing sequences is detected in orangutan, gibbon and baboon. The most parsimonious explanation for these two changes is that they are the result of the translocation of a block of sequence from 16p to 4p along the branch prior to human–chimpanzee–gorilla divergence. This translocation probably involved only a portion of the region devoted to OR sequences on 16p, since sequences derived from the 5′ end of the 3p13-contig map to 16p in the three species, and some humans carry a block of other OR sequences on 16p (7). Other significant changes that have occurred during primate evolution are the loss of cross-hybridizing sequences from chromosome 7 in gorilla, from chromosome 11 in orangutan and baboon, and from chromosome 8 in baboon. In addition, the sequence has been gained (or simply retained) on chromosome 19, and two sites on chromosome 1 in baboon.

Our findings, combined with recent observations for other subtelomeric (7,30), pericentromeric (19–21) and OR (32; V. Brand-Arpon *et al*., submitted for publication) sequences, indicate that large-scale differences among primate genomes have been grossly underestimated. The frequently quoted 2–10% sequence variation underestimates the degree of genomic variation among these species. Few of the sites detected with the multi-site clones have been stable during the divergence of primates. Our findings are consistent with the idea that copy number differences (amplification/attrition) and contextual changes (location changes) of sequences contribute to phenotypic differences among organisms and may be more significant for speciation events than subtle differences at the nucleotide level. It will be important to compare the collection of expressed OR sequences to learn how these gross structural changes affect the repertoire of olfactory receptors in the nasal epithelium or testes of these species.

The structure and distribution of OR-containing paralogs suggest a hierarchical model for the duplication process

A multi-step process of duplications and rearrangements is required to explain our observations. OR sequences and VNTR blocks appear to have duplicated to form the larger ~22 kb duplication unit, which duplicated both intra- and interchromosomally. We suspect that the large blocks of the ~63 bp repeats at the junctions of the 22 kb duplication facilitate the recombination events that lead to duplications (and deletions) of OR genes. The pGAG, Alu and Line insertions within the units may prove useful to date the duplication events and establish the relationships among some paralogs. Copies of the 22 kb duplication unit appear to be a portion of a larger region that is shared by multiple chromosomes, since clones lacking this unit still cross-hybridize to many of the same locations. The duplicated zones appear to have been further modified by rearrangements and/or insertions of relatively unique DNA. In some cases, these sequences were included in subsequent duplications (such as the MLCK unit at the 3′ end of the 3p13-contig). It is also possible that some initial duplication units were large and were later whittled down to a dense array of key functional elements [analogous to some postulated processes in intrachromosomal gene amplification (33)].

The subterminal location of many OR-containing blocks may facilitate this process, because large-scale duplications and rearrangements could occur without affecting dosage-sensitive genes in the chromosome's interior. A disproportionate number (42%) of OR-containing duplicates is located in the 10% of the genome that lies closest to the ends of chromosomes. Our observations are consistent with the notion that terminal bands could serve as a nursery for generating diversity among a subset of this gene family (7).

Large genomic duplications provide a mechanism for changing the repertoire of ORs. Functionally redundant copies could evolve through mutations to new specificities or expression patterns. Many duplicates would degenerate into pseudogenes. It is not yet clear whether the selective pressure to expand or change the OR gene repertoire is the driving force behind the large-scale duplications we observe or if OR sequences were captured inadvertently in regions with a propensity to duplicate and are innocent bystanders in the process. Sequence comparisons among the OR sequences at each location and attempts to match expressed sequences to their chromosomal location should provide some answers.

Many locations appear to contain a combination of pseudogenes and genes. We showed previously a bias for intact ORFs on chromosomes 7, 16 and 17 (6), but none of the genomic clones analyzed here maps uniquely to 16 or 7; all cross-hybridize to other locations. The duplications in common with the 11-PAC and the 3p13 contig contain apparent pseudogenes. The fact that related pseudogenes in the two regions have specific mutations in common is an indication that clusters of pseudogenes are subject to duplication. It is not known if pseudogenes serve a function; some are transcribed (34). In any case, the duplications of pseudogenes may be a negligible genomic burden when weighed against the benefits of a process that can provide new substrates for diversifying the OR repertoire.

A subset of OR-clusters is significantly diverged from other OR-containing regions

We identified a number of OR-containing clones that map to single sites in the genome. The sequences in these clones must be sufficiently diverged that they fail to cross-hybridize to other OR-containing locations. Three of the locations detected with two or more independent clones, 17p, 19p and 11p, are locations of OR genes that are already the subject of detailed genomic analysis (32,35–38). The fact that these OR clusters are among the first to be characterized in detail is probably a consequence of the fact that mapping them was uncomplicated by extensive homology with other locations in the genome. Our results also indicate that 1q21–22, 1q44, 3p14–21 and 9q32-q34 also harbor relatively unique representatives of the OR family. Our results confirm the findings of Rouquier *et al*. (6), who detected OR sequences at three of these locations. To our knowledge, our study is the first to point to 1q21–22 as a location of OR genes.

OR-locations missing from this collection of genomic clones

Several sites that are known to harbor OR genes are not represented in the collection of BACs that we have characterized here. Our strategy missed clones encompassing a block of OR genes that is duplicated near many telomeres including 3qter, 15qter and 19pter (7). Also notably lacking from the collection are BACs representing the 6p21 site of OR genes amidst the MHC gene cluster (39) and the 14q11.2 site near the TCR-α genes (6,40; GenBank accession no. U85195). Clones encompassing these OR genes do not cross-react by FISH with other OR-containing regions (C. Friedman and B.J. Trask, unpublished data). It is therefore interesting to note that two OR-BACs were identified that hybridize near the TCR-α locus on 14q11.2, but also to a second site at 15q11.2. This result lends further support to the notion that clusters of OR genes are a complex combination of single-copy sequence and segments with homology to two or more sites, reflecting an evolutionary history of duplications and rearrangements.

Fraction of the human genome devoted to the OR family

Our results suggest that large genomic duplications are responsible for the distribution, size and diversity of much of the OR gene family. Differences among primates suggest that the portion of the genome devoted to the OR family has changed during primate evolution, possibly as a mechanism for evolving the repertoire of OR genes. If we conservatively estimate that each of the paralogous regions encompasses 125 kb (and the results with the 3p13 contig suggest that the regions in some locations may be even larger), then more than ~3 Mb or 0.1% of the human genome is occupied by the duplications and unique segments that encompass members of the OR gene family. The similarity of many of these large repeats may be maintained by sequence exchanges among the copies (e.g. by unequal cross-over or gene conversion). The multiplicity of these large regions of homology poses a severe challenge to sequencing the human genome. However, the large-insert clones we identify here should prove useful for mapping the clusters in more detail in order to determine how the exquisite transcriptional control of these diverse receptors is achieved.

MATERIALS AND METHODS

Isolation of OR-containing genomic clones

A pool of OR-specific sequences was generated by PCR from genomic DNA using degenerate primers designed in the highly conserved regions of OR genes. The OR probe pool used to screen approximately four genomic equivalents of a total genomic BAC library (CITB 978SK) was generated using primers 12464 [5′-A(G/C)(A/T/C/G)TATGACCGCTATGTGGCCATCTG] and 12465 [5′-CACCACAG(A/T)CAGGTGGGA(G/T)(C/G)CACAGG] recognizing a sequence [(T/S/R)YDRYVAI] in the second cytoplasmic loop, and C(A/G)SHL(T/S)VV in TM6. The probe used to screen the chromosome 3 library (LL03NC01) was amplified using primers in the conserved domains TM2 [OR5B; PMY(F/L)FL(S/A/T/G/C)NLS] and TM7 [OR3B; M(L/F/V/I)NPF(I/M)Y(S/C)L] (8; V. Brand-Arpon *et al*., submitted for publication). Both libraries were spotted in duplicate on high-density filters. Probes were radiolabeled by random hexamer priming, and filters were washed after hybridization to a stringency of 65°C in 0.1× SSC, or in the case of the chromosome 3 filters as described by V. Brand-Arpon *et al*. (submitted for publication). BAC clones were secondarily screened by PCR using primers 12464 and 12465. PCR products of 33 of the 52 BACs that passed this secondary screen and were FISH-mapped were cloned and sequenced to confirm OR homology. A comparison of the resulting sequences will be published elsewhere (K. Chan and H. Shizuya, unpublished data). Chromosome 3 clones were confirmed to contain OR sequences by PCR using OR3B/OR5B primers or by Southern blot analyses of *Eco*RI-restricted cosmid DNAs using the OR3B/5B PCR product amplified from total genomic DNA, as described elsewhere (V. Brand-Arpon *et al*., submitted for publication). Clones were selected for FISH analyses randomly, with some bias to include representatives from the cosmid collection with different restriction enzyme patterns (V. Brand-Arpon *et al*., submitted for publication).

Additional clones were identified for FISH analyses by cosmid walking in the chromosome 3 specific library from the ends of OR-containing clones. The resulting 250 kb contig from 3p13 is described elsewhere (V. Brand-Arpon *et al*., submitted for publication). Additional clones, which may derive from any of three locations on chromosome 3, were isolated by screening the chromosome 3 library with the T3 end of cos 88 and the T7 end of PAC169 by hybridization (V. Brand-Arpon *et al*., submitted for publication). These clones were tested for OR content by PCR and/or Southern blot analyses as described above.

The LLNL-library designations (LL03NC01) for the chromosome 3 cosmids used in this study are 3-4B17, 5-4L5, 8-6A20, 11-10A14, 13-10K1, 21-19B17, 26-21N3, 28-23I22, 32-24I3, 36-24P4, 39-25D13, 40-25H11, 45-28K17, 48-31P3, 54-37B9, 57-39H23, 65-44F15, 66-45J14, 78-51I19, 79-52K19, 80-53J7, 81-53O7, 88-24I21, 96-54B7, 97-55A16, 111-18O2, 113-26H13, 117-39K8, 202-4A19, 204-4L5, 209-23E11, 215-28I23 and 219-47A22.

FISH

Metaphase spreads were prepared from PHA-stimulated peripheral blood lymphocytes of healthy humans. A few clones were additionally mapped on human lymphoblast cell cultures GM11525 and GM10977 (Coriell). Cell lines from ATCC served as representatives of chimpanzee (*Pan troglodytes*, CRL-1847), gorilla (*Gorilla gorilla*, CRL-1854), orangutan (*Pongo pygmaeus*, CRL-1850), gibbon (*Hylobates lar*, TIB-201) and sacred baboon (*Papio hamadryas*, CRL-1495).

FISH was performed as described elsewhere (41). Briefly, cosmid or BAC DNA was isolated on Qiagen tip-100 columns or an Autogen system, respectively, biotinylated by nick-translation and hybridized to metaphase spreads in the presence of excess unlabeled Cot1 DNA (added to suppress hybridization of interspersed repeats throughout the genome). The same stringency conditions were used for all experiments, i.e. hybridization at 37°C in 50% formamide, 2× SSC, 10% dextran sulfate, pH 7.0; washing for 15 min at 42°C in 3 changes of 50% formamide/2× SSC followed by 15 min at 42°C in 2× SSC. Hybridization sites were detected with two layers of avidin–FITC separated by a layer of biotinylated anti-avidin. The chromosomes were QFH-banded by DAPI staining. DAPI and FITC images were digitized separately, but in registration, using a Zeiss Axiophot microscope, a Ludl filter wheel equipped with separate DAPI and FITC excitation filters, a ChromaTechnology multiband pass emission filter, 100× NeoFluar objective, and a Princeton cooled CCD camera (Kodak 1000×1317) operated via a custom script in ScanAlytics IPLab Spectrum 3.0 software. For the image shown, the FITC image was pseudocolored red for clarity, and the DAPI banding was displayed as grey values. Hybridization signals were analyzed in at least five and more typically 10 metaphase cells per probe. Clones localizing to a single location were scored at the microscope, and only a few images were collected. For clones mapping to multiple sites, signals were scored from digitized images to avoid missing dim signals due to photobleaching during the analysis. The contrast and gain were varied interactively so that both dim and bright signals could be scored. Signal intensity was scored on a scale of 1–4, with four corresponding to a large bright signal. The average signal intensity at each location was calculated. This value measures a combination of signal intensity and labeling efficiency, because a chromosome showing no signal at a particular location was assigned the value 0.

Eleven BAC clones were blindly mapped twice using independent DNA isolates, allowing us to assess the accuracy of clone tracking and chromosomal assignments. The same result was obtained for seven duplicate pairs, including three clones with multi-site patterns. Imprecision in our assignments to 9q32–33 and 9q33–34 were identified with three duplicates. Therefore, we have combined all clones mapping to these locations into a single group. Finally, inter-sample variation in signal intensity was noted with 302B6, which produced signals at ~17 sites in one sample (as plotted in Fig. 2A), but at only the two most prominent locations, 4p16 and 11q13, in the second.

These karyotypes shown in Figures 2 and 4 were generated by encoding ISCN 450-band level ideograms (42) in PostScript, and the ideograms of Yunis and Prakesh (31) were rendered into PostScript to produce Figure 7 (G.J. van den Engh, H.F. Massa and B.J. Trask, unpublished data). These postscript files are available from the corresponding author.

PCR typing of monochromosomal somatic cell hybrid panel

The Coriell monochromosomal somatic cell hybrid panel #2 was PCR typed using the following primer pairs designed from the ends of OR-containing cosmids: the T7 end of cosmid 32 (32MF, 5′-ATCTCATGATCTGTTCCATCC; 32IR, 5′-ATTCCAGT-

TAAAGGCATAACG), the T3 end of cosmid 32 (32CF, 5′-TCCTAATATCACCGTGGCTC; 32CR, 5′-AATATGATCACAGGGTGTACC), the T3 end of cos 5 (5AF, 5′-TCTGGCCATGTGATTGCTGC; 5AR, 5′-TATAAAGCGATAACTGAGGCT), the T7 end of cos 5 (5BF, 5′-ACCGGGAACCTTTGGAATGC; 5BR, 5′-GAGAGGTAGGCCACCGGC) and the T3 end of cos 11 (11AF, 5′-TGAACCAGCTGTCCTCAGC; 11AR, 5′-TGTTCCCAAGCCACGGACC). In addition, several pairs of primers were designed in the repeat- and OR-free regions of clone ends or in the 106 kb sequenced contig from 3p13 (V. Brand-Arpon *et al*., submitted for publication) and typed on the hybrid panel. These PCR assays are denoted A–G in Figure 5. The names of the primers for assays C–F indicate their positions in the GenBank sequence entry (AF042089). [A: the T3 end of cos 88 (88AF, 5′-ATTCTCTTGCCTTGGCCTC; 88AR, 5′-TGTTTGCAGTGCCCAGCAG) marks a position ~50 kb 5′ of the OR-cluster in the 3p13-contig; B: 64AF, 5′-TATGATGTCACTTGAGACAGC, and 64AF, 5′-GACATATTCTCCATTCAATAACC, at the T3 end of clone 64); C: U22806, 5′-AGAAGAGCGAGCGGGCGACAG and L23452, 5′-CGGTGGGAGACGGGTGAGGTA; D: U45021, 5′-AGACTAGAAAAGGAGCAACACC and L45210, 5′-GTCTACAAGGAACCCCCAAAGG; E: U66952, 5′-CCAGGCCATAGAATAGTAAATA and L67132, 5′-GGCCATGACCACGGAGAACAGG; F: U91031, 5′-AGCCAGCGGAGGGATAGGTTGC and L91333, 5′-GGGGCTGAGAGGGATTGTGTGA; G: 82AF, 5′-GAATTACTCTGATGCAATGGTG and 82AR, 5′-AAACCATGTGATCTGAGCATC at the T3 end of cosmid 88.] The 25 µl PCR reactions contained 45 ng cell-line DNA, 250 µM dNTPs, 0.4 µM each primer, 1 U Perkin Elmer AmpliTAQ. Cycling conditions were 94°C for 1 min, 35 cycles of 20 s at 94°C, 30 s at 58°C (68°C for U22806/L23452 to eliminate a rodent product with the same size as the expected human product), 20 s at 72°C, followed by 5 min at 72°C.

ACKNOWLEDGEMENTS

We thank Lee Rowen for advice on the sequence analysis. This research was supported by grants from the NIH (GM57070-01 and HG01464), the US Department of Energy (DE-FG03-96ER62173), the Programme Génome du CNRS, the Fondation pour la Recherche Médicale, and NATO.

REFERENCES

1. Lancet, D. (1986) Vertebrate olfactory reception. *Annu. Rev. Neurosci*., **9**, 329–355.
2. Buck, L. and Axel, R. (1991) A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*, **65**, 175–187.
3. 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.
4. Haines, S.L. and Akeson, R.A. (1996) Chemoreceptors expressed in taste, olfactory and male reproductive tissues. *Gene*, **178**, 1–5.
5. Vanderhaeghen, P., Schurmans, S., Vassart, G. and Parmentier, M. (1997) Molecular cloning and chromosomal mapping of olfactory receptor genes expressed in the male germ line: evidence for their wide distribution in the human genome. *Biochem. Biophys. Res. Commun*., **237**, 283–287.
6. Rouquier S., Taviaux, S., Trask, B.J., Brand-Arpon, V., van den Engh, G., Demaille, J. and Giorgi, D. (1998) Distribution of olfactory receptor genes in the human genome. *Nature Genet*., **18**, 243–250.
7. 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.H., Rouquier, S., Smith, T., Wong, D.J., Youngblom, J. and van den Engh, G.J. (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.
8. Ben-Arie, N., Lancet, D., Taylor, C., Khen, M., Walker, N., Ledbetter, D.H., Carrozzo, R., Patel, K., Sheer, D., Lehrach, H. and North, M.A. (1994) Olfactory receptor gene cluster on human chromosome 17: possible duplication of an ancestral receptor repertoire. *Hum. Mol. Genet*., **3**, 229–235.
9. Mombaerts, P., Wang, F., Dulac, C., Chao, S.K., Nemes, A., Mendelsohn, M., Edmondson, J. and Axel, R. (1996) Visualizing an olfactory sensory map. *Cell*, **87**, 675–686.
10. Ngai, J., Chess, A., Dowling, M.M., Necles, N., Macagno, E.R. and Axel, R. (1993) Coding of olfactory information: topography of odorant receptor expression in the catfish olfactory epithelium. *Cell*, **72**, 667–680.
11. Ressler, K.J., Sullivan, S.L and Buck, L.B. (1993) A zonal organization of odorant receptor gene expression in the olfactory epithelium. *Cell*, **73**, 597–609.
12. Vassar, R., Ngai, J. and Axel, R. (1993) Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium. *Cell*, **75**, 309–318.
13. Chess, A., Simon, I., Cedar, H. and Axel, R. (1994) Allelic inactivation regulates olfactory receptor gene expression. *Cell*, **78**, 823–834.
14. Ressler, K.J., Sullivan, S.L. and Buck, L.B. (1994) Information coding in the olfactory system: evidence for a stereotyped and highly organized epitope map in the olfactory bulb. *Cell*, **79**, 1245–1254.
15. 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**, 884–888.
16. 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.
17. Wienberg, J., Stanyon, R., Jauch, A. and Cremer, T. (1992) Homologies in human and *Macaca fuscata* chromosomes revealed by in situ suppression hybridization with human chromosome specific DNA libraries. *Chromosoma*, **101**, 265–270.
18. Nadeau, J.H. Kosowsky, M., Steel, K.P. (1991) Comparative gene mapping, genome duplication and the genetics of hearing. *Ann. NY Acad. Sci*., **630**, 49–67.
19. Eichler, E.E., Lu, F., Shen, Y., Antonucci, R., Doggett, N.A., Moyzis, R.K., 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.
20. Eichler, E.E., Budarf, M.L., Rocci, M., Deaven, L.D., Doggett, N.K., Nelson, D.L. and Mohrenweiser, H. (1997) Interchromosomal duplications of the adrenoleukodystrophy locus: a phenomenon of pericentromeric plasticity. *Hum. Mol. Genet*., **6**, 991–1002.
21. Regnier, V., Meddeb, M., Lecointre, G., Richard, F., Duverger, A., Nguyen, V., Dutrillaux, B., Bernheim, A. and Danglot, G. (1997) Emergence and scattering of multiple neurofibromatosis (NF1)-related sequences during hominoid evolution suggest a process of pericentromeric interchromosomal transposition. *Hum. Mol. Genet*., **6**, 9–16.
22. Maresco, D.L., Chang, E., Theil, K.S., Francke, U. and Anderson, C.L. (1996) The three genes of the human FCGR1 gene family encoding Fc gamma RI flank the centromere of chromosome 1 at 1p12 and 1q21. *Cytogenet. Cell Genet*., **73**, 157–163.
23. Chen, K.S., Manian, P., Koeuth, T., Potocki, L., Zhao, Q., Chinault, A.C., Lee, C.C. and Lupski, J.R. (1997) Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. *Nature Genet*., **17**, 154–163.
24. Endo, T., Imanishi, T., Gojobori, T. and Inoko, H. (1998) Evolutionary significance of intra-genome duplications on human chromosomes. *Gene*, **205**, 19–27.
25. Kasahara, M. (1997) New insights into the genomic organization and origin of the major histocompatibility complex: role of chromosomal (genome) duplication in the emergence of the adaptive immune system. *Hereditas*, **127**, 59–65.
26. Zimonjic, D., Kelley, M., Rubin, J., Aaronson, S. and Popescu, N. (1997) Fluorescence *in situ* hybridization analysis of keratinocyte growth factor gene amplification and dispersion in evolution of great apes and humans. *Proc. Natl Acad. Sci. USA*, **94**, 11461–11465.

27. Pérez-Jurado, L.A., Wang, Y.-K., Peoples, R., Coloma, A., Cruces, J. and Francke, U. (1998) A duplicated gene in the breakpoint regions of the 7q11.23 Williams–Beuren syndrome deletion encodes the initiator binding protein TFII-I and BAP-135, a phosphorylation target of BTK. *Hum. Mol. Genet.*, **7**, 325–334.
28. Katsanis, N., Fitzgibbon, J. and Fisher, E.M.C. (1996) Paralogy mapping: identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel PBX and NOTCH loci. *Genomics*, **35**, 101–108.
29. Ritchie, R.J., Mattie, M.G. and Lalande, M. (1998) A large polymorphic repeat in the percenteomeric region of human chromosome 15q contains three partial gene duplications *Hum. Mol. Genet.*, **7**, 1253–1260.
30. Monfouilloux, S., Avet-Loiseau, H., Amarger, V., Balazs, I., Pourcel, C. and Vergnaud, G. (1998) Recent human-specific spreading of a subtelomeric domain. *Genomics*, **51**, 165–177.
31. Yunis, J.J. and Prakash, O. (1982) The origin of man: a chromosomal pictorial legacy. *Science*, **215**, 1525–1529.
32. Rouquier, S., Friedman, C., Delettre, C., van den Engh, G., Blancher, A., Crouau-Roy, B., Trask, B.J. and Giorgi, D. (1998) A gene recently inactivated in human defines a new olfactory receptor family in mammals. *Hum. Mol. Genet.*, **7**, 1337–1345.
33. Ma, C., Martin, S., Trask, B.J. and Hamlin, J.L. (1993) Sister chromatid fusion initiates amplification of the dihydrofolate reductase gene in Chinese hamster cells. *Genes Dev.*, **7**, 605–620.
34. Perry, B.N. and Connerton, I.F. (1996) Olfactory receptor-encoding genes and pseudogenes are expressed in humans. *Gene*, **16**, 247–249.
35. Glusman, G., Clifton, S., Roe, B. and Lancet, D. (1996) Sequence analysis in the olfactory receptor gene cluster on human chromosome 17: recombinatorial events affecting receptor diversity. *Genomics*, **37**, 147–160.
36. Buettner, J.A. and Evans, G.A. (1996) Organization and localization of olfactory receptor genes on human chromosome 11. *Am. J. Hum. Genet.*, **59** (suppl.), 1737.
37. Carver, E.A., Issel-Tarver, L., Rine, J., Olsen, A.S. and Stubbs, L. (1998) Location of mouse and human genes corresponding to conserved canine olfactory receptor gene subfamilies. *Mamm. Gen.*, **9**, 349–354.
38. Issel-Tarver, L. and Rine, J. (1997) The evolution of mammalian olfactory receptor genes. *Genetics*, **145**, 185–195.
39. Fan, W., Liu, Y.C., Parimoo, S. and Weissman, S.M. (1995) Olfactory receptor-like genes are located in the human major histocompatibility complex. *Genomics*, **27**, 119–123.
40. Boysen, C., Simon, M.I. and Hood, L. (1997) Analysis of the 1.1-Mb human alpha/delta T-cell receptor locus with bacterial artificial chromosome clones. *Genome Res.*, **7**, 330–338.
41. Trask, B.J. (1998) 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, Cold Spring Harbor, NY, Vol. 4, pp. 303–413.
42. ISCN (1995) In Mitelman, F. (ed.), *An International System for Human Cytogenetic Nomenclature*. S. Karger, Basel.