Genomics **71,** 296–306 (2001)
doi:10.1006/geno.2000.6431, available online at http://www.idealibrary.com on IDEAL®

Mouse–Human Orthology Relationships in an Olfactory Receptor Gene Cluster

Michal Lapidot,* Yitzhak Pilpel,*[,1] Yoav Gilad,* Ayellet Falcovitz,* Dror Sharon,*[,2] Thomas Haaf,† and Doron Lancet*[,3]

*Department of Molecular Genetics and the Crown Human Genome Center, The Weizmann Institute of Science, Rehovot 76100, Israel; and †Max-Planck-Institute of Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany

Received July 5, 2000; accepted November 2, 2000

The olfactory receptor (OR) subgenome harbors the largest known gene family in mammals, disposed in clusters on numerous chromosomes. One of the best characterized OR clusters, located at human chromosome 17p13.3, has previously been studied by us in human and in other primates, revealing a conserved set of 17 OR genes. Here, we report the identification of a syntenic OR cluster in the mouse and the partial DNA sequence of many of its OR genes. A probe for the mouse M5 gene, orthologous to one of the OR genes in the human cluster (OR17-25), was used to isolate six PAC clones, all mapping by *in situ* hybridization to mouse chromosome 11B3–11B5, a region of shared synteny with human chromosome 17p13.3. Thirteen mouse OR sequences amplified and sequenced from these PACs allowed us to construct a putative physical map of the OR gene cluster at the mouse Olfr1 locus. Several points of evidence, including a strong similarity in subfamily composition and at least four cases of gene orthology, suggest that the mouse Olfr1 and the human 17p13.3 clusters are orthologous. A detailed comparison of the OR sequences within the two clusters helps trace their independent evolutionary history in the two species. Two types of evolutionary scenarios are discerned: cases of "true orthologous genes" in which high sequence similarity suggests a shared conserved function, as opposed to instances in which orthologous genes may have undergone independent diversification in the realm of "free reign" repertoire expansion.

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. AF309122–AF309134.

[1] Present address: Department of Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115.

[2] Present address: Ocular Molecular Genetics Laboratory, Massachusetts Eye and Ear Infirmary, Harvard Medical School, 243 Charles Street, Boston, MA 02114.

[3] To whom correspondence should be addressed. Telephone: 972-8-9343683 or -934412. Fax: 972-8-9344112. E-mail: doron.lancet@weizmann.ac.il.

INTRODUCTION

Olfactory receptors (ORs) are seven-transmembrane domain proteins that underlie the recognition and G-protein-mediated transduction of odorant signals (Buck and Axel, 1991; Lancet and Pace, 1987; Mombaerts, 1999). OR genes are expressed mainly in the olfactory neuroepithelium, but were also found in other tissues (Drutel *et al.,* 1995; Walensky *et al.,* 1998) including mammalian germ cells (Parmentier *et al.,* 1992). Each olfactory sensory neuron expresses one or very few OR genes (Lancet, 1991) and probably just one allele at a given locus (Chess *et al.,* 1994). This expression pattern is believed to provide the molecular basis of odor discrimination by the sensory cells.

OR genes were first cloned in the rat (Buck and Axel, 1991) and were later found in the genomes of a wide variety of species including human (Ben-Arie *et al.,* 1994; Parmentier *et al.,* 1992; Schurmans *et al.,* 1993; Selbie *et al.,* 1992), mouse (Ressler *et al.,* 1993; Sullivan *et al.,* 1996), dog (Issel-Tarver and Rine, 1996), pig (Velten *et al.,* 1998), chicken (Nef and Nef, 1997), *Xenopus* (Freitag *et al.,* 1995), channel catfish (Ngai *et al.,* 1993), zebrafish (Barth *et al.,* 1997), opposum (Kubick *et al.,* 1997), mudpuppy (Zhou *et al.,* 1997), lamprey (Berghard and Dryer, 1998), *Caenorhabditis elegans* (Troemel *et al.,* 1995), and *Drosophila melanogaster* (Clyne *et al.,* 1999; Vosshall *et al.,* 1999). ORs are present in the genome of these species in a large germline repertoire (the "olfactory subgenome") with an estimated 500–1000 coded proteins (Buck and Axel, 1991; Lancet, 1986; Ressler *et al.,* 1994). They form an outstandingly diverse multigene family, consisting of 32 distinct families (Glusman *et al.,* 2000a; Lancet and Ben-Arie, 1993).

While some regions in the OR gene are highly conserved, others show sequence variability. Earlier analyses showed that most of the variable amino acid residues are clustered within the transmembrane helices TM3, TM4, and TM5 (Buck and Axel, 1991). More recently, an analysis of hundreds of vertebrate OR sequences, along with molecular modeling of the recep-

0888-7543/01 $35.00

tor structure, revealed a set of 17 interior-facing variable residues, which was proposed to serve as the complementarity-determining region (CDR) for odorant recognition (Pilpel and Lancet, 1999). The CDR likely accounts for the superfamily's ability to address multitudes of structurally diverse odorants.

OR genes reside in clusters dispersed throughout the genome. This organization likely reflects the evolutionary processes that led to the expansion of the OR repertoire and could play a functional role in the control of OR gene expression (Barth *et al.*, 1997; Ben-Arie *et al.*, 1994; Buettner *et al.*, 1998; Glusman *et al.*, 1996, 2000b; Issel-Tarver and Rine, 1996; Reed, 1994; Sullivan *et al.*, 1996; Troemel *et al.*, 1995; Vanderhaeghen *et al.*, 1997). The best characterized genomic OR cluster resides on human chromosome 17p13.3, encompassing ~450 kb and containing 17 OR coding regions, 6 of which are pseodogenes (Ben-Arie *et al.*, 1994; Glusman *et al.*, 1996, 2000b). These genes belong to six subfamilies (1D, 1E, 1G, 1P, 1R, and 3A). The extensive knowledge regarding this cluster makes it an ideal target for comparative studies of OR evolution. The orthologous cluster was also reported in a number of nonhuman primates (Sharon *et al.*, 1999), revealing an overall conservation. In the present study, we attempted to characterize the corresponding cluster in the mouse genome to shed light on its long-term evolution.

In the mouse genome, 12 OR clusters were mapped by genetic linkage analysis to loci on seven different chromosomes (chromosomes 1, 2, 7, 9, 10, 11, and 13) (Copeland *et al.*, 1993; Sullivan *et al.*, 1996). Additional clusters were identified in later studies (Asai *et al.*, 1996; Carver *et al.*, 1998; Strotmann *et al.*, 1999; Szpirer *et al.*, 1997). The availability of a large number of murine clusters made it likely that one or more human–mouse orthologous cluster pairs could be identified. This assumption is based on the notion that at least some of the clusters formed prior to the human–mouse divergence from their common ancestor (Glusman *et al.*, 2000b). However, the definition of orthologous pairs for clusters and individual genes is not always straightforward, because of species-specific duplication events. The present study describes the resolution of some of these problems, leading to the identification of mouse orthology relationships for the human chromosome 17p13.3 cluster as well as for some of its constituent genes. The results provide insight relevant to the evolution and function of the OR repertoire.

MATERIALS AND METHODS

PCR and primers. Primers for PCR amplification and for sequencing were synthesized according to previous publications. Human OR degenerate primers were designed according to Ben-Arie *et al.* (1994), and M5 primers were designed according to a published sequence (Sullivan *et al.*, 1996). Novel primers were designed using Oligo-Primer Analysis Software, version 5.1, by Wojciech and Piotr Rychlik (NBI), and Amplify Software for PCR, version 2.53b, created by Bill Engels (Department of Genetics, University of Wisconsin, Madison).

PCRs were performed in a total volume of 25 μl, containing a 0.2 mM concentration of each deoxynucleotide (Promega, Madison, WI), 50 pmol of each primer, (1 ml of 50 pmol/ml), PCR buffer containing 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris, pH 8.3, 1 unit of *Taq* DNA polymerase (Boehringer Mannheim, Mannheim, Germany), and 50 ng genomic DNA or 10 ng PAC DNA. PCR conditions were as follows: (1) For OR5B, OR5A, OR3A, and OR3B, there were 35 cycles of 1 min at 94°C, 1 min at 55°C, 1 min at 72°C. The first step of denaturation and the last step of extension were each 3 min long. (2) For M5 5′ and M5 3′ (M5 probe preparation), conditions were the same as those listed above. (3) For RH mapping, each OR gene was amplified under unique conditions. For OR17-24 and OR17-25, there were 35 cycles of 30 s at 94°C, 30 s at 55°C, 30 s at 72°C. The first step of denaturation and the last step of extension were each 3 min long. For OR17-40, there were 9 cycles of 1 min at 94°C, 1 min at 68°C, 1 min at 72°C, followed by 20 cycles of 1 min at 94°C, 1 min at 64°C, 3 min at 72°C. The first step of denaturation and the last step of extension were 3 and 10 min long, respectively. For OR17-210, there were 10 cycles of 1 min at 94°C, 1 min at 60°C, 1 min at 72°C followed by 30 cycles of 1 min at 94°C, 1 min at 50°C, 2 min at 72°C. The first denaturation step was 1 min, and the final elongation step was 3 min.

RH mapping. The Stanford G3 RH01 RH mapping panel (Research Genetics, Inc., Huntsville, AL) was screened by PCR with four gene-specific primer sets, corresponding to OR17-24, OR17-25, OR17-40, and OR17-210. The products were analyzed by electrophoresis on 1% agarose gels, and scores were submitted electronically for analysis at the Stanford Human Genome Center.

Isolation of M5-positive PACs from a mouse genomic DNA library. Specific primers designed according to the published sequence of M5 (GenBank Accession No. U28780) were used to amplify this sequence from mouse genomic DNA. The product (336 bp) was extracted from an agarose gel (1.5%) using Qiagen's Qiaquick kit, radiolabeled, and used as a probe for screening the RCPI21 Female (129S6/SvEvTac) Mouse PAC library (http://bacpac.med.buffalo.edu). The PAC library was constructed by Kazutoyo Osoegawa and Pieter de Jong (Roswell Park Cancer Institute). The screening process was carried out at the Resource Center of the German Human Genome Project (the Max-Planck-Institute for Molecular Genetics). Positively hybridizing PAC clones were received as stabs in agar and were immediately streaked out and grown overnight at 37°C on kanamycin plates. Single colonies were produced and tested for OR content by PCR. The primer pairs used were different combinations of the OR degenerate primers OR5B, OR3B, OR5A, and OR3A. These primers corresponding to trans-membranal helices 2, 3, 6, and 7, respectively, were designed according to rat OR cDNA conserved regions (Buck and Axel, 1991). M5 primers were employed to assess the presence of M5 in the various PACs.

Fluorescence in situ hybridization (FISH). FISH analysis was performed through a collaboration with Dr. Thomas Haaf (the Max-Planck-Institute of Molecular Genetics, Berlin, Germany). Chromosomes were prepared from the Moloney murine leukemia virus-transformed cell line WMP-1, derived from wild mice of the strain WMP/WMP. WMP-1 cells carry pairs of metacentric Robertsonian (Rb) translocation chromosomes that are morphologically distinguishable and, therefore, greatly facilitate physical mapping in the mouse (Zoernig *et al.*, 1995). For FISH (Ward *et al.*, 1995), chromosome preparations were treated with 100 μg/ml RNase A in 2× SSC at 37°C for 60 min and with 0.01% pepsin in 10 mM HCl at 37°C for 10 min and then dehydrated in an ethanol series (70, 85, and 100%). Slides were denatured at 80°C in 70% formamide, 2× SSC, pH 7.0, and again dehydrated in an alcohol series. PAC DNA was labeled by standard nick-translation with biotin-16–dUTP (Boehringer Mannheim). Biotinylated PAC DNA (10 ng/μl) was coprecipitated with 100 ng/μl mouse cot-1 competitor DNA (Gibco) and 500 ng/μl salmon sperm carrier DNA and redissolved in 50% formamide, 10% dextran sulfate, 2× SSC. After 10 min of denaturation at 70°C, 30 μl of hybridization mixture was applied to each slide and sealed under a

coverslip. Slides were left to hybridize in a moist chamber at 37°C for 1 to 3 days. Slides were washed 3× 5 min in 50% formamide, 2× SSC at 42°C followed by a 5-min wash in 0.1× SSC at 65°C. Hybridized probes were detected by fluorescein isothiocyanate (FITC)-conjugated avidin (Vector). Chromosomes and cell nuclei were counterstained with 1 μg/ml 4,6-diamidino-2-phenylindole (DAPI) in 2× SSC for 5 min. The slides were mounted in 90% glycerol, 100 mM Tris–HCl, pH 8.0, and 2.3% DABCO. Images were taken with a Zeiss epifluorescence microscope equipped with a thermoelectronically cooled charge-coupled device camera (Photometrics CH250), which was controlled by an Apple Macintosh computer. Oncor imaging software was used to capture grayscale images and to superimpose the images into a color image. Oncor imaging software was also used to invert the DAPI image into a G-banded metaphase for identification of the chromosomes.

The PACs used are RCPIP711J2199 (PAC2), RCPIP711H04134 (PAC3), RCPIP711D16225 (PAC4), RCPIP711M06287, (PAC5), RCPIP711K15373 (PAC6), and RCPIP711F21384 (PAC7) from the mouse PAC library RPCI21.

Cloning and sequencing of OR coding regions. PAC DNA was extracted using a Qiagen plasmid kit (Qiagen, Chatworth, CA), according to the manufacturer's recommendations for very-low-copy-number plasmids. PCR was performed on DNA of individual PACs. Primers were modified for subsequent subcloning into the pAMP1 vector. The products were subcloned into the pAMP1 vector, without prior purification, using the Clone Amp System (Gibco BRL). DNA of subclones was extracted using Wizard Plus SV minipreps DNA purification system (Promega) and sequenced using vector primers from both directions. Sequencing was performed on a Model 373A or 377 automatic DNA sequencer (PE Applied Biosystems Inc., Foster City, CA), using a DyeDeoxy terminator cycle sequencing kit and AmpliTaq DNA polymerase FS (Perkin–Elmer, Foster City, CA).

Sequence analysis. Sequencing reactions were performed on PCR products or clones in both directions. Base-calling was performed using the ABI Analysis Software (version 3.0), and the analyzed data were edited using the Sequencher program (GeneCodes Corp., Version 3.0).

Since we did not obtain full coding regions for the genes, the partial open reading frames were conceptually translated using FASTY (Pearson *et al.*, 1997) by assembly to a "core" of properly translated OR gene sequences. Corrupted open reading frames (putative pseodogenes) bearing frameshift mutations were corrected by the assembly to the core in a fashion that allows their proper alignment with intact OR genes.

Identity for pairwise comparisons was calculated using the GeneAssist program (PE Applied Biosystems). Multiple sequence alignments and neighbor-joining analysis were performed using ClustalX (Higgins *et al.*, 1996), with standard parameters. Confidence was estimated using 1000 runs of bootstrapping. Phylogenetic trees were generated using TreeView software (Page, 1996). The ratio of synonymous and nonsynonymous substitutions per site between pairs of orthologous sequences was calculated using the subroutine Diverge from the GCG package. Family assignments and nomenclature are derived from a scheme of olfactory receptor gene classification (Glusman *et al.*, 2000a).

GenBank accession numbers. The new sequences described in this work are as follows: mOR11-4 (AF309122), mOR11-208 (P) (AF309123), mOR11-40a (AF309124), mOR11-40b (AF309125), mOR11-25 (AF309126), mOR11-7a (AF309127), mOR11-7b (AF309128), mOR11-7c(P) (AF309129), mOR11-2a (AF309130), mOR11-2b(P) (AF309131), mOR11-2c (AF309132), mOR11-2d(P) (AF309133), mOR11-2e (AF309134). Additional OR sequences used in this study are as follows: OR17-1 (AF087915), OR17-2 (AF087916), OR17-4 (AF087917), OR17-6 (AF155225), OR17-7 (AF087918), OR17-23 (AF087919), OR17-24 (AF087920), OR17-25 (AF087921), OR17-30 (AF087922), OR17-31 (AF087923), OR17-40 (AF087924), OR17-93 (AF087925), OR17-201 (AF087926), OR17-208 (AF087927), OR17-209 (AF087928), OR17-210 (AF087929), OR17-228 (AF087930), M5 (U28780), MMTPCR35P (X89687), MMTPCR50P (X89688), RATOLFPROQ (M64391), *Mus musculus* OR H3 (AF102538), CFDTMT (X64996).

Databases. Sequences and mapping information were retrieved from the following databases: (1) Genome Database (GDB), Johns Hopkins University School of Medicine (Baltimore, MD) (http://gdbwww.gdb.org); (2) Mouse Genome Database (MGD), Mouse Genome Informatics, The Jackson Laboratory (Bar Harbor, ME) (http://www.informatics.jax.org/); (3) The Unified Database (UDB) (Chalifa-Caspi *et al.*, 1997) (http://bioinformatics.weizmann.ac.il/udb); and (4) HORDE, Human or Data Exploratorium (http://bioinformatics.weizmann.ac.il/HORDE/).

RESULTS

Syntenic OR Clusters

To identify orthologous OR genes, we examined which of the previously known OR gene clusters in the human and mouse genomes are included within regions of conserved synteny. Using the relevant coordinates in the MGD linkage map (http://www.Informatics.jax.org), each of the published mouse OR clusters (Carver *et al.*, 1998; Strotmann *et al.*, 1999; Sullivan *et al.*, 1996; Szpirer *et al.*, 1997) was associated with its most likely human shared synteny region (cf. legend to Fig. 1). The exact assignment was made based on flanking genes, which have been mapped in both species. The human genomic region was then searched for the existence of a known OR cluster (Rouquier *et al.*, 1998; Fuchs *et al.*, 2001), leading to the identification of 16 candidate shared synteny clusters.

We further searched for a case in which a pair of known OR coding sequences, residing within syntenic clusters, share over 85% amino acid sequence identity, suggesting true gene orthology. This level of identity (85.0 $\pm$ 0.4% at the DNA and protein levels) was set as a cutoff based on a comparative analysis of 1196 orthologous mouse and human sequences (Makalowski *et al.*, 1996). Only one relevant gene pair was found by this procedure. This was M5, a 336-bp sequence residing on the mouse Olfr1 cluster on chromosome 11 (Sullivan *et al.*, 1996), which shares 87% amino acid identity with OR17-25 in the human 17p13.3 OR cluster. Importantly, the human OR cluster happens to be most thoroughly characterized, including a complete DNA sequence (Ben-Arie *et al.*, 1994; Glusman *et al.*, 1996, 2000b).

Fine-Mapping of the Human OR Cluster

While the human OR cluster has previously been mapped by FISH to the cytogenetic band 17p13.3 just centromeric to the Miller–Dieker syndrome (MDS) critical region (Ben-Arie *et al.*, 1994; Kwiatkowski *et al.*, 1990), a more accurate localization was deemed necessary to ensure the shared synteny relationship to Olfr1. Four gene-specific primer sets, corresponding to the coding regions of OR17-24, OR17-25, OR17-40, and OR17-210, were mapped onto the Radiation Hybrid map of the Stanford Human Genome Center. ORs 17-24 and 17-25 showed radiation hybrid linkage to

marker D17S1798, and ORs 17-40 and 17-210 showed linkage to marker D17S1828. The STS marker D17S1548 (4.52 UDB Mb) was found by electronic PCR to reside on the telomeric sequence of the cluster (Glusman *et al.*, 2000b). The relationships to other genes and markers are provided by the UDB integrated map of chromosome 17 (http://bioinformatics.weizmann.ac.il/udb) (Chalifa-Caspi *et al.*, 1997).

Mouse Genomic Clone Isolation

Based on the notion that M5 is the likely orthologue of OR17-25, we used the M5 sequence to identify genomic clones potentially spanning the mouse Olfr1 cluster. The M5 coding region segment was resequenced from mouse gDNA using end primers designed according to the flanks of the published 336-bp sequence. Sequence analysis of independent PCR products from three individual mice showed a difference from the published sequence by only two adjacent bases, possibly reflecting a previous sequencing error or a polymorphism. The M5 PCR product was radiolabeled and used as a probe for screening a mouse P1-derived artificial chromosome (PAC) library (RCPI21 mouse PAC). Six positively hybridizing PAC clones were obtained, three of which (PAC2, PAC3, and PAC6) showed a weaker hybridization signal (Fig. 1a).

The presence of OR coding sequences in all six PAC clones was confirmed by PCR using different combinations of degenerate primers OR5B, OR3B, OR5A, and OR3A designed according to four conserved regions within the OR coding region (Ben-Arie *et al.*, 1994). For all six PAC clones, these amplification experiments gave the combination of four products with the expected molecular masses, serving as an OR signature (data not shown). In parallel, the presence of the specific OR coding region sequence M5 was verified by PCR amplification using M5-specific primers. Only four of the PAC clones (PAC4, PAC5, PAC6, and PAC7) gave a positive signal. The other two (PAC2 and PAC3) correspond to two of the clones that hybridized weakly with the M5 probe. It is likely that the PAC2 and PAC3 clones were selected in the genomic screen due to hybridization with the highly similar receptor sequences mOR11-40a and mOR11-40b, which belong to the same subfamily (3A).

To verify that all six PAC clones arise from a single OR cluster presumably located on mouse chromosome 11, a FISH analysis was performed. The PAC clones were hybridized individually to metaphase mouse chromosome spreads. All six PACs showed specific hybridization to mouse chromosome 11B3–11B5 (Fig. 1a). Under routine conditions of high stringency, no hybridization signals were visible on other chromosomes. The results are in agreement with the expected location of the mouse cluster by synteny relationship with the human 17p13.3 cluster and with the genetic mapping of Olfr1 (Sullivan *et al.*, 1996).

Sequencing of OR Coding Regions

Each PAC clone underwent PCR amplification with tailed OR-specific OR5B/OR3B primers and was subcloned into a high-yield vector. Minilibraries were thus created for each PAC, from which multiple OR subclones were sequenced using vector primers. The risk of PCR recombinants (Ben-Arie *et al.*, 1994; Glusman *et al.*, 1996) was maximally reduced by regarding only sequences obtained from two independent PCR-based libraries (for PAC2, PAC4, and PAC5) or from libraries of two genomic clones that share their OR content (for PAC3 and PAC6). PAC7 contained only sequences that were already known from PAC4 and served as an additional confirmation for their validity.

Thirteen different OR sequences were identified (Fig. 1b), 10 of which were completely novel. Two of the sequences were almost identical to previously published sequences: mOR11-2a differed by a single base from MMTPCR50P, and the pseudogene mOR11-2b was different at only 1 bp relative to the seemingly intact MMTPCR35P, both isolated from mouse testis cDNA (Vanderhaeghen *et al.*, 1997). Neither one of these genes has been previously assigned a chromosomal localization. The sequences reported here are highly dependable, since each was identically found in three to five minilibraries and in more than 20 subclones (Fig. 1b). The difference with respect to the previously published sequences may constitute an experimental error or a polymorphism. Cases in which one gene has both an intact and a nonfunctional allele are known to exist in the OR superfamily (Sharon *et al.*, 1999).

The third previously known OR is mOR11-25, which contains a segment almost identical to mouse M5 (Sullivan *et al.*, 1996), which served here as a probe for the mouse cluster. The M5 gene failed to be amplified by the OR5B/OR3B primers and therefore could not be subcloned and sequenced in the same fashion as the other ORs. The known partial sequence (336 bp) was extended in the 5′ direction using randomly primed PCR, obtaining a sequence of 789 bp that reached 63 bp upstream of the first ATG.

The OR sequences residing on each PAC served to construct a putative content-based contig map that includes PAC3, PAC4, PAC5, PAC6, and PAC7. PAC2 does not share any OR sequences with the other PAC clones and was thus assigned a presumed position based on the orthology with the human cluster (see below).

Orthology Relations with the Human OR Cluster

Each of the sequences in the mouse OR cluster was conceptually translated and was analyzed for sequence similarity against 224 human OR sequences (Fuchs *et al.*, 2001). This led to a significant result: for all the OR genes from mouse chromosome 11 reported here, the closest human sequence was from the OR cluster on

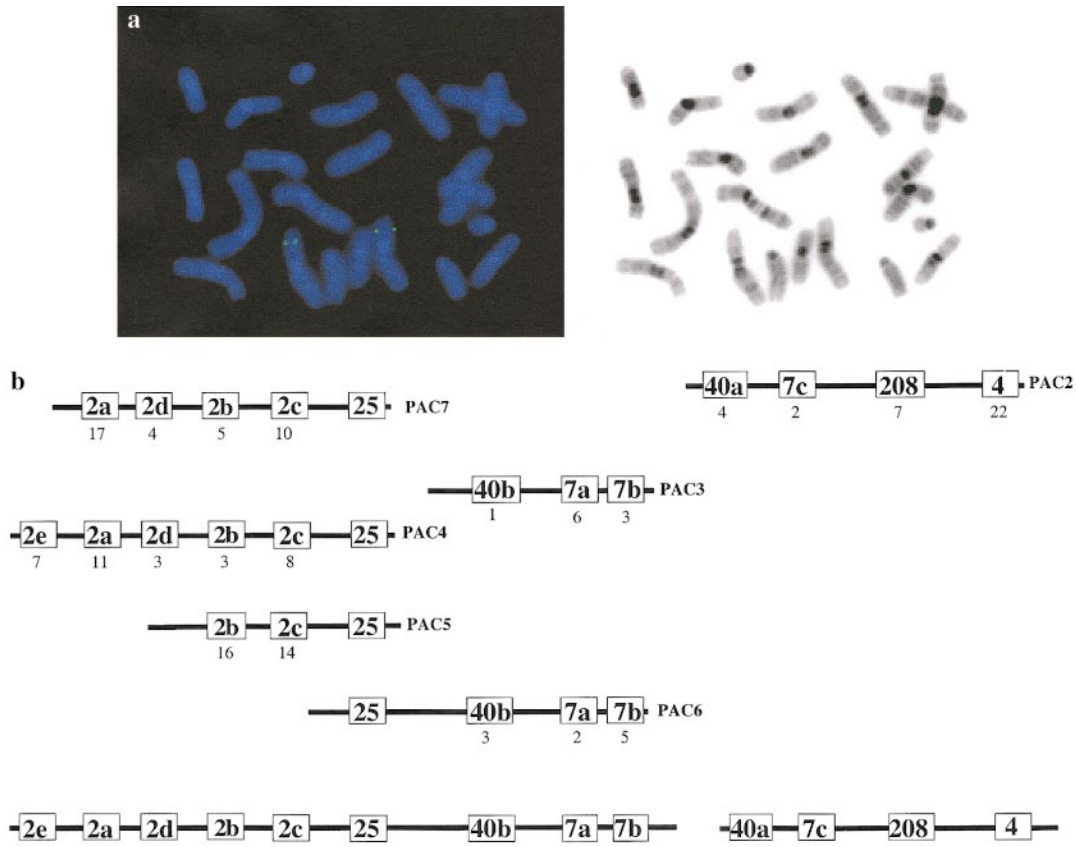

FIG. 1. (**a**) Representative fluorescence *in situ* hybridization for PAC2, one of the five genomic clones that cover the mouse OR cluster. Metaphase spreads of the permanent suspension cell line WMP-1 were probed with biotinylated DNA and were detected by FITC-avidin (green fluorescence). Chromosomes were counterstained with DAPI (blue). All PACs mapped to the same mouse chromosomal region, 11B5 of the WMP-1 marker Rb(1.11)2Mp1. This position corresponds to the mouse Olfr1 cluster on chromosome 11, 44 cM (Sullivan *et al.*, 1996), and is clearly distinct from all other OR clusters with shared synteny and potential orthology in human (Ben-Arie *et al.*, 1994; Buettner *et al.*, 1998; Fan *et al.*, 1995; Rouquier *et al.*, 1998; Trask *et al.*, 1998; Volik *et al.*, 1995) and mouse (Carver *et al.*, 1998; Strotmann *et al.*, 1999; Sullivan *et al.*, 1996; Szpirer *et al.*, 1997). Synteny relations were derived from the Mouse Genome Database linkage maps (http://www.informatics.jax.org/). The other specific cluster pairs are: mouse Olfr2 (olfr39) at Chr 9, 5–6 cM, syntenic to human 19p13.1–p13.2; mouse Olfr3 at Chr 2, 24.6 cM, syntenic to human 9q32–q34; mouse Olfr4 (olfr4-1, olfr4-2) at Chr 2, 52–53 cM, syntenic to human 11q11–q13; mouse Olfr5 at Chr 7, 0.5 cM, syntenic to human 11p15; mouse Olfr6 at Chr 7, 49.5 cM, syntenic to human 11q13; mouse Olfr7 at Chr 9, 23 cM, syntenic to human 11q24; mouse Olfr8 (olfr40) at Chr 10, 41.5–44 cM, syntenic to human 19p13.1–p13.2; mouse Olfr9 at Chr 10, 69.5 cM, syntenic to human 12q13; mouse Olfr10 at Chr 11, 30 cM, syntenic to human 5q34; mouse Olfr11 at Chr 13, 9 cM, syntenic to human 6p21 and to human 7p15 (the cluster is on the border between two regions of conserved synteny each bearing a human OR cluster); mouse Olfr12 at Chr 1, 53 cM, syntenic to human 2q22–q23; mouse Olfr16 at Chr 1, 94.2 cM, syntenic to human 1q21–q23; mouse Olfr37 (Olfr37a–Olfr37e) at Chr 4, 21.5 cM, syntenic to human 9q22; Olfr38 at Chr 6, 22.5 cM, syntenic to human 7q35; Olfr89 at Chr 17, 20–21 cM, syntenic to human 6p21. (**b**) Tentative physical map of the PAC clones covering the mouse OR gene cluster (lowest row). The map was constructed based on open reading frame content and on homology to the fully sequenced human cluster. There is one breakpoint between PAC2 and all the rest. The numbers below each sequence denote the number of OR subclones sequenced for a given coding region. Differences in the counts probably relate to priming preferences. The novel mouse sequences were assigned locus-related trivial names based on sequence similarity to the respective human genes. When more than one mouse gene was related to the same human gene, the mouse genes were labeled by consecutive lowercase letters.

human chromosome 17p13.3. The range of similarity values for these closest pairs was 74–88% identity at the protein level (Table 1 and Fig. 2). There was also a clear correspondence in subfamily content: all subfamilies found in the mouse cluster (3A, 1A, 1D, 1E, and 1P) were also present in the human cluster. Only two subfamilies found in the human cluster, each represented by one gene (subfamilies 1G and 1R), were absent from the mouse cluster. This strongly suggested that the mouse and human clusters are orthologous. In addition to their human counterparts, some of the mouse sequences showed a high degree of identity (>80%) to sequences in other mammalian species. mOR11-208 is 93.22% identical to the mouse sequence AF102538 (Krautwurst *et al.*, 1998), mOR11-2c is 81.48% identical to the canine sequence DTMT (Parmentier *et al.*, 1992), and mOR11-2e is 89.81% identical to the rat sequence RATOLFPROQ (Buck and Axel, 1991).

Further inspection revealed a somewhat complex picture (Fig. 2): while in a few cases simple pairwise orthology was seen, in other cases multiple potential mouse orthologous genes were found to exist for a single human sequence, and vice versa. In yet other instances, no obvious mutual orthology was found, as judged using a liberal cutoff of 82% identity

TABLE 1

Identity Scores for Mouse OR vs Human Sequences

Subfamily 3A	17-24 (P)	17-25 (P)	17-40	17-201	17-228
MMOR11-40a	68.2 (74.3)	64.8 (70.1)	87.5 (85.5)	75.9 (79.8)	79.6 (80.9)
MMOR11-40b	69.2 (74.4)	65.2 (71.1)	85.7 (84.6)	76.4 (80.1)	79.2 (80.6)
MMOR11-25	67.8 (71.4)	84.3 (85.8)	67.4 (70.7)	69.8 (73.5)	67.4 (72.1)
Subfamily 1A			17-6		17-7
MMOR11-7a			80.4 (82.0)		86.5 (83.4)
MMOR11-7b			80.8 (82.2)		86.0 (83.4)
MMOR11-7c (P)			73.4 (79.7)		77.6 (80.1)
Subfamily 1D	17-4	17-23 (P)	17-30		17-31
MMOR11-4	87.4 (85.6)	84.2 (8.4)	79.5 (81.1)		80.00 (81.1)
Subfamily 1E		17-2	17-93		17-210 (P)
MMOR11-2a		74.5 (78.7)	72.9 (78.4)		67.4 (72.9)
MMOR11-2b (P)		80.1 (81.7)	78.2 (82.1)		72.1 (76.3)
MMOR11-2c		81.5 (81.0)	77.8 (79.3)		73.5 (77.1)
MMOR11-2d (P)		79.2 (83.1)	76.9 (81.2)		71.2 (76.7)
MMOR11-2e		75.5 (77.8)	71.6 (76.7)		68.4 (75.2)
Subfamily 1P					17-208 (P)
MMOR11-208 (P)					76.6 (81.6)
AF102538					89.9 (82.8)

Note. Each mouse sequence was subjected to pairwise comparisons with all human sequences from the cluster that belong to the same subfamily. Numbers are protein sequence identity, and the numbers in parentheses indicate nucleotide sequence identity. Pseodogenes are denoted by (P). Values were calculated using GeneAssist software. The mouse sequence AF102538 (Krautwurst *et al.,* 1998), of unknown genomic localization, was included in the table, although not recovered in this study, as its high sequence identity (89.9%) to OR 17-208 suggests potential orthology.

(Makalowski *et al.,* 1996). In these cases, mouse and human OR sequences still belonged to the same subfamily, but appeared to have diverged significantly from one another.

To examine further the orthology relationships, we computed the ratio of synonymous to nonsynonymous nucleotide substitutions (K_s/K_a) for every human–mouse pair belonging to a given subfamily. A correlation diagram was drawn for the K_s/K_a values vs the amino acid identity scores (Fig. 3). Three main groups were revealed: (1) K_s/K_a values higher than 6.0 and percentage identities higher than 85%. This group likely constitutes true orthologous genes: (2) K_s/K_a values between 2.5 and 6.0 and percentage identities ranging between 65 and 82%. These probably represent nonorthologous receptor pairs that still show considerable evolutionary preservation: (3) K_s/K_a values of 1.2 or less, with percentage identities in the same range as group 2. All the receptor pairs in this group belong to subfamily 1E. These receptors seem to have undergone a species-specific process of diversification.

True OR orthologous genes are expected to share a function and therefore to display higher conservation at the residues that constitute the odorant binding site. Figure 4 shows a correlation between the interorthologue variability and the interparalogue variability. The variability value for paralogous genes (V_p, abscissa) was computed from a multiple alignment of 197 ORs as described (Pilpel and Lancet, 1999). The variability value for orthologous genes (V_o, ordinate) was calculated from an alignment of the six orthologous pairs by summing variability profiles (Pilpel and Lancet, 1999) of individual orthologous pairs. Low variability in both paralogous genes and orthologous genes (lower left quadrant) indicates conserved positions. High values for both suggest randomly disposed variability (upper right quadrant). When orthologous residue pairs show high variability but paralogous genes are more conserved, this may indicate species-specific functional sites shared by many gene family members (upper left quadrant). Finally, low interorthologue variability but high diversity among paralogous genes may indicate residues at which conservation holds across species but variation among gene family members is high, potentially related to functional repertoire diversity (lower right quadrant). Importantly, all but 1 of the 17 residues identified as constituting the complementarity determining regions of the OR protein (Pilpel and Lancet, 1999) appear in this quadrant. This serves to confirm that the orthologous pairs are correctly assigned.

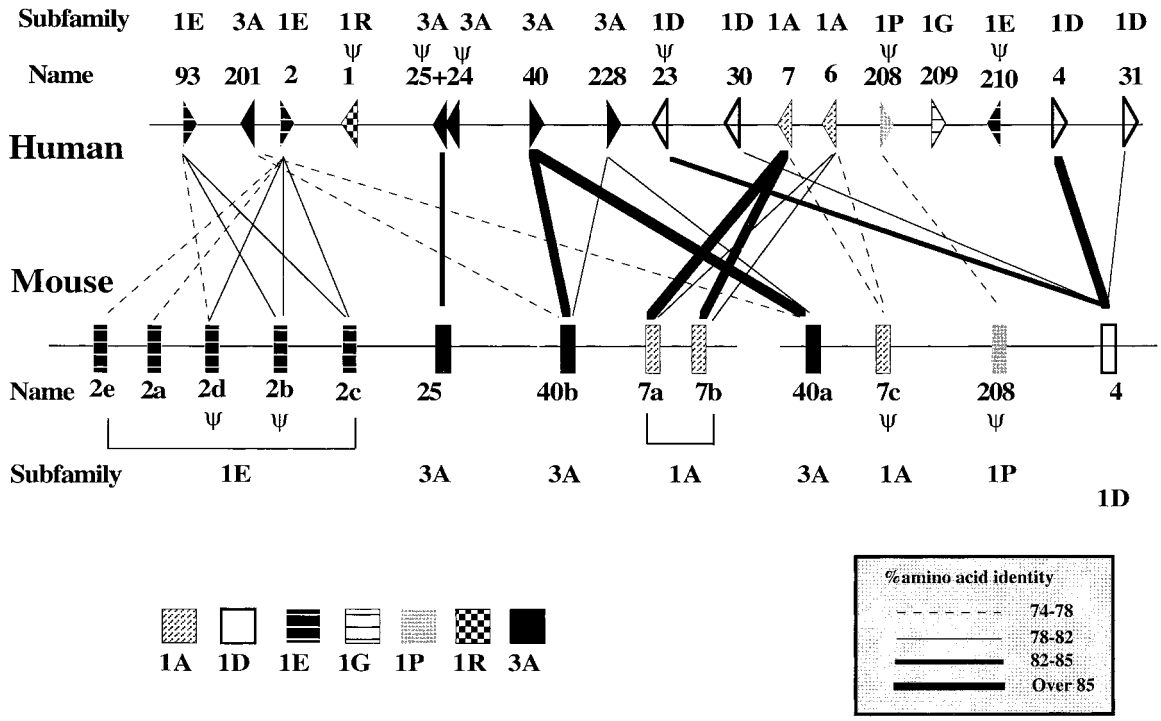

FIG. 2. The human and mouse clusters are displayed one on top of the other. Genes belonging to the same subfamily appear in the same pattern. Related human and mouse sequences within each subfamily (i.e., genes sharing over 74% amino acid sequence identity) are connected by a line. The width and pattern of the connecting line are correlated to the degree of similarity between the pair of related sequences. Ψ indicates a pseudogene. The top line shows the nomenclature symbols, e.g. 1E is OR1E.

DISCUSSION

A Unique OR Cluster

The genomic clones that contain OR genes within the mouse Olfr1 cluster have been isolated through probing with one partial mouse OR sequence, M5. The mouse genome contains an estimated 1000 rather similar OR genes, disposed in several dozen clusters. It could therefore be expected that an individual OR probe might identify numerous genomic clones on multiple chromosomal loci (Carver *et al.*, 1998; Trask *et al.*, 1998). Still, under the stringency conditions used, only six PAC clones turned out to be M5-positive, and they all localized to the same region on mouse chromosome 11, suggesting that they cover a single genomic cluster.

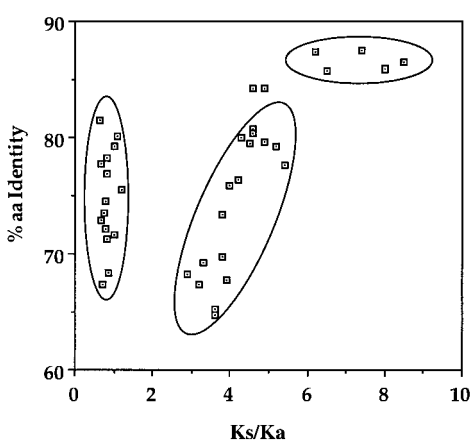

FIG. 3. The ratio of synonymous to nonsynonymous nucleotide substitutions (K_s/K_a) is plotted against the percentage amino acid identity. This is performed for each pair of human and mouse genes belonging to the same subfamily (Table 1). Ovals indicate correlation groups. K_s/K_a values were calculated using the Diverge subroutine of the GCG package.

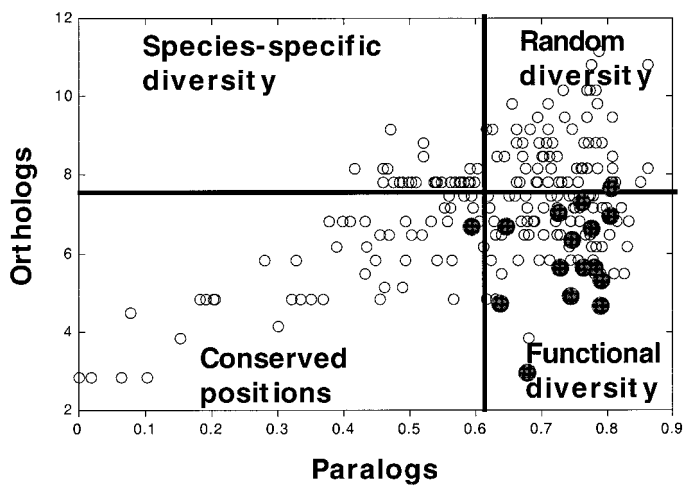

FIG. 4. A "variability diagnostic plane" analysis, applied to six human–mouse orthologous genes (human 17-4 with mouse 11-4; human 17-7 with mouse 11-7a; human 17-7 with mouse 11-7b; human 17-7 with mouse 11-7c; human 17-40 with mouse 11-40a; and human 17-40 with mouse 11-40b). A correlation is shown between the interorthologue variability (V_o) and the interparalogue variability (V_p) for all amino acid positions. The lower right quadrant represents residues that have high variability among paralogous genes, but relatively low variability among orthologous genes. The 17 CDR residues are shown as dark circles.

Interestingly, only four of the six genomic clones contained the M5 gene, as scored by PCR. The other two clones must have hybridized to the M5 probe via a cross-reaction with the mOR11-40a and mOR11-40b genes, which belong to the same subfamily (3A). It may be inferred that OR genes belonging to subfamily 3A are not very prevalent in the mouse genome and are found chiefly at the Olfr1 locus. Indeed, an analysis of 224 human OR genes (Fuchs *et al.*, 2001) has revealed only 6 members of subfamily 3A, 5 of which reside within the syntenic chromosome 17p13.3 cluster. The success of the present study largely hinges on the uniqueness of subfamily 3A in the OR gene repertoire, in contrast with more widespread OR subfamilies, e.g., subfamily 7E (Fuchs *et al.*, 2001).

The clear single locus hybridization of all six PAC clones corresponding to the Olfr1 cluster suggests that it is unique not only in containing a rather rare OR subfamily, but also in terms of its overall sequence features. This is in distinction to the recently reported overall homology among a set of ~20 OR loci, distributed on 13 human chromosomes. In a genome-wide study, only a small set of OR locations have been reported to behave as single-copy (Trask *et al.*, 1998). This subset indeed included the OR cluster on human chromosome 17p13.3, in line with the results presented here.

Primate-Specific Events

The comparison of a mouse and a human OR cluster is instrumental in shedding light on some events that may have occurred late in evolution, on the primate branch. The human OR cluster accommodates a fused gene pair, OR17-24 and OR17-25. The similarity between these two sequences and their codirectionality suggests that they arose by a tandem duplication and were later fused to each other. OR17-25 is clearly a pseudogene, due to deletion of two bases, and it is likely that OR17-24 has also lost its function after the gene fusion event (Glusman *et al.*, 1996).

M5 is presumably the mouse orthologue of OR17-25. Our sequencing by elongation of the mOR11-25 coding region shows that it does not bear the 2-bp deletion that rendered the human gene nonfunctional. It also indicates that the upstream region, deleted in human, is intact in mouse. It is thus likely that the gene fusion event found in human has not occurred in mouse.

Subfamily 1P has one member in the human cluster (OR17-208). Our work revealed one homologous sequence on the mouse cluster (mOR11-208), but its similarity is not high enough to be an orthologue. Interestingly, a potential mouse orthologue with a very high identity score (89.9%) has recently been published (Krautwurst *et al.*, 1998), but its genomic location remains unknown. This OR gene may reside in an as yet uncharacterized region of the currently studied cluster. Alternatively, it may be present on a different cluster, in which case its translocation mechanism would need to be determined.

The fraction of OR pseudogenes in the human genome is estimated as higher than 50% (Buettner *et al.*, 1998; Fuchs *et al.*, 2001; Rouquier *et al.*, 1998), while the published mouse sequences appear to consist of only a few percent pseudogenes. In the currently studied OR cluster, the pseudogene count is 6/16 (38%) for human, compared to 4/13 (31%) for mouse. The difference is not statistically significant. Yet, the process of pseudogene formation appears to have taken place independently in the two mammalian species, as relatively little overlap exists between the two sets of inactive OR genes. Only one pseudogene is shared in common, OR-208, and it is defective in different ways in the two species. Interestingly, the mouse counterparts for two intact human genes constitute mixed groups of genes and pseudogenes: mOR11-7c is a mouse pseudogene with two intact paralogous genes (7a and 7b). mOR11-2b and 2d are pseudogenes, with three intact paralogous genes (2a, 2c, and 2e). It is thus possible that such mouse pseudogenes were formed relatively recently, following an extensive duplication process.

OR Cluster Evolution

As part of our screen for orthologous OR clusters, we have assembled a comprehensive table of syntenic OR clusters in human and mouse (Fig. 1, legend). In 12 of the mouse OR loci (Strotmann *et al.*, 1999; Sullivan *et al.*, 1996; Szpirer *et al.*, 1997), the shared human synteny has not been pointed out previously. For the other 4 clusters, on mouse chromosomes 2, 6, 9, and 10, shared synteny has previous been reported (Carver *et al.*, 1998; Issel-Tarver and Rine, 1997). Furthermore, in each of these cases, the syntenic cluster pair was shown to contain OR genes belonging to the same subfamily. It appears that a majority of the mouse OR clusters identified to date have a syntenic counterpart in human. This suggests that genomic identity and locations of OR gene clusters date back as far as the dawn of mammals. This would suggest an early process of cluster nucleation (Glusman *et al.*, 2000b).

In a previous comparative study of the OR locus on human chromosome 17p13.3, it was revealed that a conservation pattern applies also to individual genes within the cluster. Thus, practically every human gene had a distinct orthologue in the various simian clusters (Sharon *et al.*, 1999). This was consistent with the notion that the cluster has undergone few species-specific rearrangements or segmental duplications in the past 15 million years. In contrast, a comparison of the human and the mouse OR clusters displayed a more complex evolutionary history.

Based on the notion that true human–mouse orthology calls for unique gene pairing with ~85% amino acid sequence identity, there may be relatively few such pairs in the OR cluster studied here. It should

also be stressed that any final assignment of orthologous pairs should await the availability of sequences for the entire OR repertoire in both mouse and human. Our results suggest a scenario whereby a minimal ancient cluster has undergone an independent process of internal gene duplication and deletion in each of the species. Some human genes have two or more mouse homologues and vice versa. In other cases, a comparison within a given subfamily reveals several members in each species, with similarity scores too low for true orthology. Thus, it appears that in the period since mouse and human diverged from each other, there has been a continuous process of genome dynamics, with gene duplication events independently taking place in the two evolutionary branches. This is consistent with our dendogram analysis (Fuchs *et al.*, 2001) showing a continuous OR repertoire expansion in the past 80–100 million years. It should be pointed out, however, that this applies to some OR clusters, while others could reveal higher levels of internal conservation.

OR Functional Evolution

It is interesting to compare the percentage of amino acid identity shared between the paralogous genes OR17-40 and OR17-228 (81.21%) to the percentage identity between OR17-40 and its mouse orthologue mOR11-40a (87.50%). The human and mouse sequences diverged ~80 million years ago. Assuming that the duplication that created OR17-40 and OR17-228 occurred shortly thereafter, the difference in conservation between each gene pair reflects a difference in selection level between classical orthologous and paralogous genes. The same amount of neutral mutations occurred in all genes, but while the pair of orthologues was subject to conservative selection, OR17-228 was free to diverge from its paralogue.

The comparison of the mouse and human clusters reveals an apparent dichotomy between two principal cases: (1) human genes with distinct mouse orthologues sharing ~85% amino acid sequence identity and (2) subfamilies containing a few members in each species, but with no well-defined orthologues. This may reflect the existence of two evolutionary modes within the olfactory receptor repertoire, leading to the appearance of "generalist" and "specialist" receptors (Fig. 5).

Although the comparison of individual clusters provides important information, certain evolutionary questions can be addressed only by a genome-wide approach. For instance, evaluation of the size and of the pseudogene fraction in the mouse OR repertoire can be obtained by amplification of OR sequences from the entire mouse genome. This approach complements similar efforts in our laboratory, which are currently being carried out to elucidate the human olfactory subgenome.

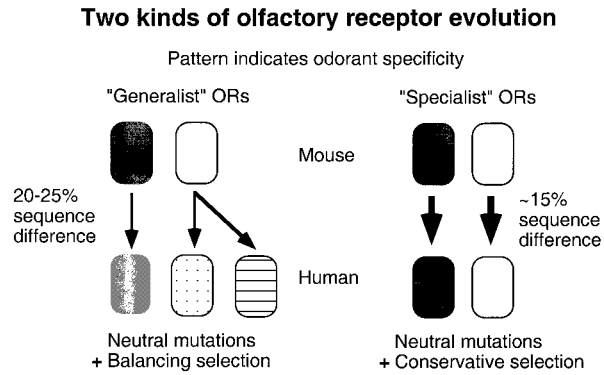

FIG. 5. "Specialist" ORs are receptors that have evolved for a distinct odorant bind, potentially common to many different mammals. The proteins encoded by these genes may odorants whose recognition is crucial for survival, such as poisonous compounds. The binding site of specialist ORs may behave much like that of single-copy genes, which tend to be conserved among different species. "Generalist" ORs are receptors that do not subserve a conserved chemosensory function and are thus free to diversify. Groups of these receptors may have evolved separately in different species to cope with the various odorants encountered. Generalist OR genes have no equivalent in the world of "one gene–one phenotype." There is a low level of selection against the loss of single generalist OR genes, as no gene is significant when standing alone. The main selective force is maintaining the size and diversity of the entire repertoire. Along these lines, some of the mammalian OR gene repertoire has most likely evolved in a free-reign regimen, with a possible contribution of balancing selection, known to generate diversity for diversity's sake. This may have ensured the creation of a repertoire that is large and eclectic enough to be fully functional. Specialist ORs are subject to sequence-conserving purifying selection, as may be indicated by their manifesting a high overall amino acid sequence identity score (~85%) between human and mouse. Generalist ORs may be subject to positive selection, and their binding sites may tend to undergo increased diversity for diversity's sake (Gilad *et al.*, 2000). This may be seen through somewhat lower identity scores (~75–80%).

ACKNOWLEDGMENTS

Doron Lancet holds the Ralph and Lois Silver Chair in Human Genomics This work was supported by the Crown Human Genome Center, by a Ministry of Science grant to the National Laboratory for Genome Infrastructure, by the National Institutes of Health (DC00305), by the Krupp foundation, by the German–Israeli Foundation for Scientific Research and Development, and by the Weizmann Institute Glasberg, Levy, Nathan, Brunschwig, and Levine funds.

REFERENCES

Asai, H., Kasai, H., Matsuda, Y., Yamazaki, N., Nagawa, F., Sakano, H., and Tsuboi, A. (1996). Genomic structure and transcription of a murine odorant receptor gene: Differential initiation of transcription in the olfactory and testicular cells. *Biochem. Biophys. Res. Commun.* **221:** 240–247.

Barth, A. L., Dugas, J. C., and Ngai, J. (1997). Noncoordinate expression of odorant receptor genes tightly linked in the zebrafish genome. *Neuron* **19:** 359–369.

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.

Berghard, A., and Dryer, L. (1998). A novel family of ancient vertebrate odorant receptors. *J. Neurobiol.* **37:** 383–392.

Buck, L., and Axel, R. (1991). A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell* **65:** 175–187.

Buettner, J. A., Glusman, G., Ben-Arie, N., Ramos, P., Lancet, D., and Evans, G. A. Organisation and evolution of olfactory receptor genes on human chromosome 11. *Genomics.* **53:** 56–68.

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. Genome* **9:** 349–354.

Chalifa-Caspi, V., Rebhan, M., Prilusky, J., and Lancet, D. (1997). The Unified DataBase (UDB): A novel genome integration concept. *HUGO Genome Digest* **4:** 15.

Chess, A., Simon, I., Cedar, H., and Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. *Cell* **78:** 823–834.

Clyne, P., Warr, C., Freeman, M., Lessing, D., Kim, J., and Carlson, J. (1999). A novel family of divergent seven-transmembrane proteins: Candidate odorant receptors in *Drosophila. Neuron* **22:** 327–338.

Copeland, N. G., Jenkins, N. A., Gilbert, D. J., Eppig, J. T., Maltais, L. J., Miller, J. C., Dietrich, W. F., Weaver, A., Lincoln, S. E., Steen, R. J., Stein, L. D., Nadeau, J. H., and Lander, E. S. (1993). A genetic linkage map of the mouse: Current applications and future prospects. *Science* **262:** 57–66.

Drutel, G., Arrang, J. M., Diaz, J., Wisnewsky, C., Schwartz, K., and Schwartz, J. C. (1995). Cloning of OL1, a putative olfactory receptor and its expression in the developing rat heart. *Receptors Channels* **3:** 33–40.

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.

Freitag, J., Krieger, J., Strotmann, J., and Breer, H. (1995). Two classes of olfactory receptors in *Xenopus laevis. Neuron* **15:** 1383–1392.

Fuchs, T., Glusman, G., Horn-Saban, S., Lancet, D., and Pilpel, P. (2001). The human olfactory subgenome: from sequence to structure and evolution. *Human Genetics* In press.

Gilad, Y., Segre, D., Sharon, D., Skorecki, K., Nachman, W. M., and Lancet, D. (2000). Dichotomy of single- nucleotide polymorphism haplotypes in olfactory receptors genes and pseudogenes. *Nature Genetics* **26:** 221–224.

Glusman, G., Bahar, A., Sharon, D., Pilpel, Y., White, J., and Lancet, D. (2000a). The olfactory receptor gene superfamily: Data mining, classification and nomenclature. *Mamm. Genome.* **11:** 1016–1023.

Glusman, G., Clifton, S., Roe, R., 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.

Glusman, G., Sosinsky, A., Ben-Asher, E., Avidan, N., Sonkin, D., Bahar, A., Rosenthal, A., Clifton, S., Roe, B., Ferraz, C., Demaille, J., and Lancet, D. (2000b). Sequence, structure and evolution of complete human olfactory receptor gene cluster. *Genomics* **63:** 227–245.

Higgins, D. G., Thompson, J. D., and Gibson, T. J. (1996). Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* **266:** 383–402.

Issel-Tarver, L., and Rine, J. (1996). Organization and expression of canine olfactory receptor genes. *Proc. Natl. Acad. Sci. USA* **93:** 10897–10902.

Issel-Tarver, L., and Rine, J. (1997). The evolution of mammalian olfactory receptor genes. *Genetics* **145:** 185–195.

Krautwurst, D., Yau, K. W., and Reed, R. R. (1998). Identification of ligands for olfactory receptors by functional expression of a receptor library. *Cell* **97:** 917–926.

Kubick, S., Strotmann, J., Andreini, I., and Breer, H. (1997). Subfamily of olfactory receptors characterized by unique structural features and expression patterns. *J. Neurochem.* **69:** 465–475.

Kwiatkowski, D. J., Akolog, L., Ledbetter, D. H., and Morton, C. C. (1990). Identification of the functional profilin gene, its localization to chromosome subband 17p13.3, and demonstration of its deletion in some patients with Miller–Dieker syndrome. *Am. J. Genet.* **46:** 559–567.

Lancet, D. (1986). Vertebrate olfactory reception. *Annu. Rev. Neurosci.* **9:** 329–355.

Lancet, D. (1991). Olfaction. The strong scent of success [News]. *Nature* **351:** 275–276.

Lancet, D., and Ben-Arie, N. (1993). Olfactory receptors. *Curr. Biol.* **3:** 668–674.

Lancet, D., and Pace, U. (1987). The molecular basis of odor recognition. *Trends Biochem. Sci.* **12:** 63–66.

Makalowski, W., Zhang, J., and Boguski, M. S. (1996). Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences. *Genome Res.* **6:** 846–857.

Mombaerts, P. (1999). Seven-transmembrane proteins as odorant and chemosensory receptors. *Science* **286:** 707–711.

Nef, S., and Nef, P. (1997). Olfaction: Transient expression of a putative odorant receptor in the avian notochord. *Proc. Natl. Acad. Sci. USA* **94:** 4766–4771.

Ngai, J., Dowling, M. M., Buck, L., Axel, R., and Chess, A. (1993). The family of genes encoding odorant receptors in the channel catfish. *Cell* **72:** 657–666.

Page, R. D. (1996). TreeView: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12:** 357–358.

Parmentier, M., Libert, F., Schurmans, S., Schiffmann, S., Lefort, A., Eggerickx, D., Ledent, C., Mollereau, C., Gerard, C., Perret, J., *et al.* (1992). Expression of members of the putative olfactory receptor gene family in mammalian germ cells. *Nature* **355:** 453–455.

Pearson, W. R., Wood, T., Zhang, Z., and Miller, W. (1997). Comparison of DNA sequences with protein sequences. *Genomics* **46:** 24–36.

Pilpel, T., and Lancet, D. (1999). The variable and conserved functional sites of olfactory receptor proteins. *Protein Sci.* **8:** 969–977.

Reed, R. R. (1994). The molecular basis of sensitivity and specificity in olfaction. *Semin. Cell Biol.* **5:** 33–38.

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.

Ressler, K. J., Sullivan, S. L., and Buck, L. B. (1994). A molecular dissection of spatial patterning in the olfactory system. *Curr. Opin. Neurobiol.* **4:** 588–596.

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. *Nat. Genet.* **18:** 243–250.

Schurmans, S., Muscatelli, F., Miot, F., Mattei, M. G., Vassart, G., and Parmentier, M. (1993). The OLFR1 gene encoding the HGMP07E putative olfactory receptor maps to the 17p13 → p12 region of the human genome and reveals an MspI restriction fragment length polymorphism. *Cytogenet. Cell. Genet.* **63:** 200–204.

Selbie, L. A., Townsend, N. A., Iismaa, T. P., and Shine, J. (1992). Novel, G protein-coupled receptors: A gene family of putative human olfactory receptor sequences. *Mol. Brain Res.* **13:** 159–163.

Sharon, D., Glusman, G., Pilpel, Y., Khen, M., Gruetzner, F., Haaf, T., and Lancet, D. (1999). Primate evolution of an olfactory receptor cluster: Diversification by gene conversion and recent emergence of pseodogenes. *Genomics* **61:** 24–36.

Strotmann, J., Hoppe, R., Conzelmann, S., Feinstein, P., Momberts, P., and Breer, H. (1999). Small subfamily of olfactory receptor genes: Structural features, expression pattern and genomic organization. *Gene* **236:** 281–291.

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.

Szpirer, C., Szpirer, J., Riviere, M., Tazi, R., and Pontarotti, P. (1997). Mapping of the Olf89 and Rfp genes to the rat genome: Comparison with the mouse and human and new insights into the evolution of the rodent genome. *Cytogenet. Cell. Genet.* **78:** 137–139.

Trask, B. J., Massa, H., Brand-Apron, V., Chan, K., Friedman, C., Nguyen, O. T., Eichler, E., Van den Engh, G., Rouquier, S., Shizuya, H., and Giorgi, D. (1998). Large multi-chromosomal duplications encompass many members of the olfactory receptor gene family in the human genome. *Hum. Mol. Genet.* **7:** 2007–2020.

Troemel, E. R., Chou, J. H., Dwyer, N. D., Colbert, H. A., and Bargmann, C. I. (1995). Divergent seven transmembrane receptors are candidate chemosensory receptors in *C. elegans*. *Cell* **83:** 207–218.

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.

Velten, F., Rogel-Gaillard, C., Renard, C., Pontarotti, P., Tazi-Ahnini, R., Vaiman, M., and Chardon, P. (1998). A first map of the major histocompatibility complex class I region. *Tissue Antigens* **51:** 183–194.

Volik, S., Lebedev, Y., Nikolaev, L., Schevchenko, Y., Vinogradova, T., Kopantzev, E., Kolesnik, T., Monastyrskaya, G., Kunz, U., and Grzeschik, K. H. (1995). Mapping of transcribed sequences on human chr. 19. *DNA Seq.* **6:** 13–26.

Vosshall, L., Amrein, H., Morozov, P., Rzhetzky, A., and Axel, R. (1999). A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* **96:** 725–736.

Walensky, L. D., Ruat, M., Bakin, R. E., Blackshaw, S., Ronnett, G. V., and Snyder, S. H. (1998). Two novel odorant receptor families expressed in spermatids undergo 5′-splicing. *J. Biol. Chem.* **273:** 9378–9387.

Ward, D. C., Boyle, A., and Haaf, T. (1995). Fluorescence *in situ* hybridization techniques. Metaphase chromosomes, interphase nuclei, and extended chromatin fibers. *In* "Human Chromosomes: Principles and Techniques" (R. S. Verma and A. Babu, Eds.), pp. 184–192, MacGraw-Hill, New York.

Zhou, Q., Hinkle, G., Sogin, M., and Dionne, V. (1997). Phylogenetic analysis of olfactory receptor genes from mudpuppy. *Biol. Bull.* **193:** 248–250.

Zornig, M., Klett, C., Lovec, H., Hameister, H., Winking, H., Adolph, S., and Moroy, T. (1995). Establishment of permanent wild-mouse cell lines with readily identifiable marker chromosomes. *Cytogenet. Cell. Genet.* **71:** 37–40.