MHC-Linked Olfactory Receptor Loci Exhibit Polymorphism and Contribute to Extended HLA/OR-Haplotypes

Anke Ehlers,[1] Stephan Beck,[2] Simon A. Forbes,[3] John Trowsdale,[3] Armin Volz,[1] Ruth Younger,[2] and Andreas Ziegler[1,4]

[1]*Institut für Immungenetik, Universitätsklinikum Charité, Humboldt-Universität zu Berlin, 14050 Berlin, Germany;* [2]*The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK;* [3]*Department of Pathology, Division of Immunology, University of Cambridge, Cambridge CB2 1QP, UK*

Clusters of olfactory receptor (OR) genes are found on most human chromosomes. They are one of the largest mammalian multigene families. Here, we report a systematic study of polymorphism of OR genes belonging to the largest fully sequenced OR cluster. The cluster contains 36 OR genes, of which two belong to the vomeronasal 1 (V1-OR) family. The cluster is divided into a major and a minor region at the telomeric end of the HLA complex on chromosome 6. These OR genes could be involved in MHC-related mate preferences. The polymorphism screen was carried out with 13 genes from the HLA-linked OR cluster and three genes from chromosomes 7, 17, and 19 as controls. Ten human cell lines, representing 18 different chromosome 6s, were analyzed. They were from various ethnic origins and exhibited different HLA haplotypes. All OR genes tested, including those not linked to the HLA complex, were polymorphic. These polymorphisms were dispersed along the coding region and resulted in up to seven alleles for a given OR gene. Three polymorphisms resulted either in stop codons (genes *hs6M1–4P*, *hs6M1–17*) or in a 16–bp deletion (gene *hs6M1–19P*), possibly leading to lack of ligand recognition by the respective receptors in the cell line donors. In total, 13 HLA-linked OR haplotypes could be defined. Therefore, allelic variation appears to be a general feature of human OR genes.

[The sequence data reported in this paper have been submitted to EMBL under accession nos. AC006137, AC004178, AJ132194, AL022727, AL031983, AL035402, AL035542, Z98744, CAB55431, AL050339, AL035402, AL096770, AL133267, AL121944, Z98745, AL021808, and AL021807.]

Studies on inbred mice and rats have shown that products of MHC genes influence individual-specific odors (Yamazaki et al. 1979, 1999; Singh et al. 1987; Brown et al. 1989; Penn and Potts 1998a; for review, see Penn and Potts 1998b), which play an important role in mate choice (Yamazaki et al. 1976; Potts et al. 1991; Penn and Potts 1998c): MHC-dissimilar mating partners are preferred. It has even been demonstrated that the reproductive performance of mice can be influenced via urine odors by a discrete point mutation in the H-2K gene (Yamazaki et al. 1986). A preference for MHC-dissimilar odor types has also been reported for humans: Odors were rated the more attractive the fewer HLA-class I antigens were shared by the provider of the odor and its recipient (Wedekind et al. 1995; Wedekind and Füri 1997). Although currently unproven, an interaction between the products of MHC genes or molecules associated with them and linked OR loci, as suggested by Yamazaki and colleagues (1976), could have evolved either to avoid the disadvantages connected with inbreeding or to favor the selection of an MHC-divergent partner leading to MHC-heterozygous offspring with improved protection against attack by parasites (Potts and Wakeland 1993; Beauchamp and Yamazaki 1997; Penn and Potts 1998b). Linked genes controlling mating preferences are not uncommon in other phyla: For example, in the mating and sexual development of the mushroom *Coprinus cinereus*, several closely linked polymorphic pheromone and pheromone receptor genes determine B mating-type specificities (O'Shea et al. 1998).

Vertebrates have evolved two chemosensory systems that are able to discriminiate a large array of scents: the main olfactory epithelium (MOE) and the vomeronasal organ (VNO). The MOE apears to discriminate odors from the environment (conscious odor perception; Zhao et al. 1998), whereas receptors expressed in the VNO appear to recognize substances such as pheromones that result in behavioral responses that do not involve higher cognitive centers of the brain (subconscious odor perception; Wysocki 1989; Leinders-Zufall et al. 2000). The OR of the VNO belong

[4]Corresponding author.
E-MAIL andreas.ziegler@charite.de; FAX 49-30-450-53953.
Article and publication are at www.genome.org/cgi/doi/10.1101/gr.120400.

10:1968–1978 ©2000 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/00 $5.00; www.genome.org

to two distinct families (V1-OR and V2-OR) of ~50–100 genes each (Dulac and Axel 1995; Herrada and Dulac 1997; Matsunami and Buck 1997), while the family of the unrelated MOE-expressed OR (M-OR) genes contains up to 1000 members in mammals, including man (Buck and Axel 1991). The three types of polypeptides are G-protein-coupled receptors (GPCR) with putative seven transmembrane domains (TM1–TM7); they comprise also extracellular (EC1–EC4) and cytoplasmatic regions (CP1–CP4).

In humans, M-OR genes are located on nearly all chromosomes (Rouquier et al. 1998b), and a few M-OR loci in close linkage to the *HLA-F* locus were described several years ago (Fan et al. 1995; Gruen et al. 1996). We have recently demonstrated that at least 36 OR genes, two of them of the V1-OR type and the rest M-OR loci, are located in the immediate vicinity of the *HLA-F* locus in one major and one minor cluster (Younger et al. 2000; Ziegler et al. 2000a). Alleles of these genes would be subject to the strong linkage disequilibrium that is a characteristic feature of this chromosomal region (Malfroy et al. 1997; Tay et al. 1997; Naruse et al. 1998). So far, allelic variations of human OR genes have not been described (Mombaerts 1999a,b), although such variations might be expected to contribute to individual odor perception. They could be particularly pronounced for OR genes in close linkage to the highly polymorphic HLA class I loci.

To provide a foundation for further studies, this work aims to determine whether OR genes, and in particular those in HLA-linkage, exhibit polymorphisms. To this end, we have concentrated on the potentially expressible M-OR genes with open reading frames (ORF) and investigated whether allelic variation is a regular feature of these loci. Thirteen HLA-linked and, for comparison, three OR genes from chromosomes 7, 17, and 19 were analyzed in detail using 10 cell lines from different ethnic origins and with different HLA haplotypes. The results indicate that all M-OR loci analyzed are polymorphic.

RESULTS

Genomic Organization of HLA-Linked OR Genes

The MHC-linked OR clusters contain at least 36 OR loci (a detailed account of the genomic organization will be published elsewhere [Younger et al. 2000]), 34 of which are members of the M-OR family, and two of which are of the V1-OR type (Fig. 1). From sequencing of PACs and BACs covering the region telomeric of *HLA-F*, 15 of the identified M-OR genes (*hs6M1-1*, *-3*, *-6*, *-10*, *-12*, *-15*, *-16*, *-17*, *-18*, *-20*, *-21*, *-27*, *-28*, *-32*, *-35*) showed complete open reading frames (ORFs) of the expected length and, therefore, were predicted to be functional. The remaining 19 loci (*hs6M1-2P*, *-4P*, *-5P*, *-7P*, *-8P*, *-9P*, *-13P*, *-14P*, *-19P*, *-22P*, *-23P*, *-24P*, *-25P*, *-26P*, *-29P*, *-30P*, *-31P*, *-33P*, *-34P*) appeared to be pseudogenes. With the exception of the *hs6M1-10*, *-32*, and *-35* genes, all other OR loci with intact ORF were located in the major HLA-linked OR cluster between *HLA-F* and *RFP*. The following 22 M-OR genes were analyzed in detail by sequencing: 21 are located within the 560 kb directly telomeric of the *GABBR1* locus (Peters et al. 1998; Younger et al. 2000; *hs6M1-1*; *-2P*, *-3*, *-4P*, *-5P*, *-6*, *-7P*, *-8P*, *-12*, *-13P*, *-15*, *-16*, *-17*, *-18*, *-19P*, *-20*, *-21*, *-23P*, *-24P*, *-25P*, *-26P*), whereas *hs6M1-10* is part of the minor HLA-linked OR cluster distal of the *RFP* gene (Fig. 1). To compare the degree of polymorphism between HLA-linked and non-HLA-linked OR genes, three additional OR genes (*hs7M1-1*, *hs17M1-20*, and *hs19M1-4*) located on chromosomes 7, 17, and 19, respectively, were investigated. Sequences were determined for all OR genes in 10 cell lines (Table 1) chosen from different ethnic groups and selected to possess distinct HLA class I haplotypes to maximize the chance of detecting polymorphisms in HLA-linked OR loci. The high degree of homology exhibited by some of the OR genes—for example, *hs6M1-3* and *hs6M1-6* or *hs6M1-12* and *hs6M1-16* (Younger et al. 2000; Ziegler et al. 2000b)—was taken into account during the design of primers for amplifications by PCR.

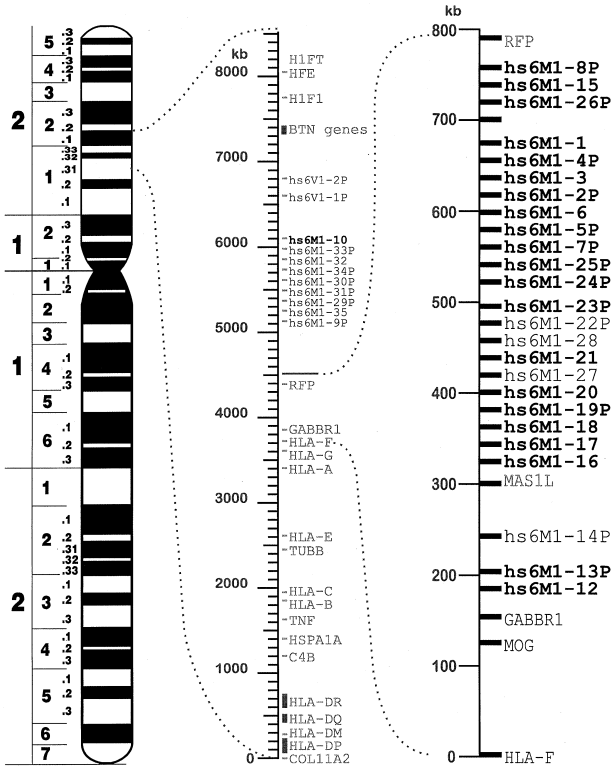

Figure 1 Chromosome 6 ideogram. All identified OR genes in the vicinity of the HLA complex (*hs6M1-1* to *-35P*, and the V1-OR type loci *hs6V1-1P*, *-2P*) and their approximate location on the physical map are indicated. The genes highlighted in bold type are those that have been analyzed here in more detail.

Table 1. Nucleotide and Amino Acid Substitutions and Mutations

Domain, amino acid position	Cell line, HLA type, ethnic origin										PAC/BAC
	BM28.7	BM19.7	LG2	KR3598	H2LCL	WT51	SA	YAR	OLGA	AMAI	
	A1, B35	A2, B13	A2, B27	A2, B44	A3, B7	A23, B65	A24, B7	A26, B38	A31, B62	A68, B53	
	Black	Black	Caucasian	Caucasian	Caucasian	Caucasian	Japanese	Jewish	Am. Indian	Algerian	
hs6M1-1											
TM3 107	Leu CTA	Leu CTG	Leu CTA	Leu CTA	Leu CTA	Leu CTA	Leu CTA	Leu CTA	Leu CTA	Leu CTA	Leu CTA
	*01	*02	*01	*01	*01	*01	*01	*01	*01	*01	*01
	Allele-designation										
hs6M1-3											
TM3 113	Ala GCA	Ala GCA	Thr ACA	Ala GCA	Thr ACA	Thr ACA	Thr Ala ACA GCA	Thr ACA	Thr Ala ACA GCA	Thr ACA	Thr ACA
CP3 226	Gln CAA	Gln CAA	Arg CGA	Gln CAA	Arg CGA	Arg CGA	Arg Gln CGA CAA	Arg CGA	Arg Gln CGA CAA	Arg CGA	Arg CGA
CP3 228	Val GTA	Val GTA	Val GTA	Val GTA	Val GTA	Ile ATA	Val GTA	Val GTA	Val GTA	Val GTA	Val GTA
TM6 261	Ile ATA	Ile ATA	Ile ATA	Ile ATA	Ile ATA	Met ATG	Ile ATA	Ile ATA	Ile ATA	Met ATG	Ile ATA
	*02	*02	*01	*02	*01	*03	*01 *02	*01	*01 *02	*04	*01
hs6M1-4P											
EC1 14	Ile ATT	Ile ATT	Ile ATT	Ile ATT	Ile ATT	Leu CTT	Ile ATT	Ile ATT	Ile ATT	Leu CTT	Ile ATT
EC2 84	Val GTG	Val GTG	Val GTG	Val GTG	Val GTG	Val GTG	Val GTG	Val GTG	Val GTG	Val GTC	Val GTG
EC2 99	Thr ACG	Thr ACG	Thr ACA	Thr ACA	Thr ACA	Thr ACG	Thr ACA	Thr ACA	Thr ACA	Thr ACG	Thr ACA
EC3 182	Val GTT	Val GTT	Ala GCT	Ala Val GCT GTT	Ala GCT	Val GTT	Ala GCT	Ala GCT	Ala Val GCT GTT	Val GTT	Ala GCT
EC3 194	Gln CAG	Gln CAG	STOP TAG	STOP Gln TAG CAG	STOP TAG	Gln CAG	STOP TAG	STOP TAG	STOP Gln TAG CAG	Gln CAG	STOP TAG
TM5 206	Ile ATT	Ile ATT	Ile ATT	Ile ATT	Ile ATT	Ile ATA	Ile ATT	Ile ATT	Ile ATT	Ile ATA	Ile ATT
	*05	*05	*01	*01 *02	*01	*03	*01	*01	*01 *02	*04	*01
hs6M1-6											
TM2 74	Tyr TAC	Tyr TAC	His CAC	Tyr His TAC CAC	His CAC	Tyr TAC	Tyr His TAC CAC	His CAC	Tyr His TAC CAC	Tyr TAC	His CAC
TM3 111	Ala GCA	Ala GCA	Ala GCA	Ala GCA	Ala GCA	Thr ACA	Ala GCA	Ala GCA	Ala GCA	Ala GCA	Ala GCA
TM3 120	Ser TCG	Ser TCG	Ser TCG	Ser TCG	Ser TCG	Ser TCA	Ser TCG	Ser TCG	Ser TCG	Ser TCG	Ser TCG
TM4 146	Val GTT	Val GTT	Ala GCT	Val Ala GTT GCT	Ala GCT	Val GTT	Val Ala GTT GCT	Ala GCT	Val Ala GTT GCT	Val GTT	Ala GCT
TM5 214	Leu CTC	Leu CTC	Leu CTG	Leu Leu CTC CTG	Leu CTG	Leu CTC	Leu Leu CTC CTG	Leu CTG	Leu Leu CTC CTG	Leu CTC	Leu CTG
TM5 218	Thr ACC	Thr ACC	Ala GCC	Thr Ala ACC GCC	Ala GCC	Thr ACC	Thr Ala ACC GCC	Ala GCC	Thr Ala ACC GCC	Thr ACC	Ala GCC
	*02	*02	*01	*02 *01	*01	*03	*02 *01	*01	*02 *01	*02	*01
hs6M1-10											
CP3 234	Gln CAA	Gln CAA	Gln CAA	Arg CGA	Gln CAA	Gln CAA	Gln CAA	Gln CAA	Gln CAA	Gln CAA	Gln CAA
	*01	*01	*01	*02	*01	*01	*01	*01	*01	*01	*01
hs6M1-12											
EC1 19	Pro CCA	Pro CCA	Pro CCA	Pro CCA	Pro CCA	Pro Pro CCA CCG	Pro CCA	Pro CCA	Pro Pro CCA CCG	Pro CCA	Pro CCA
TM1 30	Phe TTC	Phe TTC	Leu CTC	Phe Leu TTC CTC	Leu CTC	Phe TTC	Phe Leu TTC CTC	Leu CTC	Phe TTC	Phe TTC	Phe TTC
TM1 37	Leu CTA	Leu CTA	Leu CTA	Leu CTA	Leu CTA	Leu Leu CTA CTG	Leu CTA	Leu CTA	Leu Leu CTA CTG	Leu CTA	Leu CTA
TM1 48	Ala GCG	Val GTG	Ala GCG	Val Ala GTG GCG	Ala GCG	Ala Val GCG GTG	Val Ala GTG GCG	Ala GCG	Ala Val GCG GTG	Ala GCG	Ala GCG
EC2 78	Gln CAA	Gln CAA	Gln CAA	Gln CAA	Gln CAA	Gln Gln CAA CAG	Gln CAA	Gln CAA	Gln Gln CAA CAG	Gln CAA	Gln CAA
	*01	*02	*03	*02$_a$ *03$_a$	*03	*01$_a$ *04$_a$	*02$_a$ *03$_a$	*03	*01$_a$ *04$_a$	*01	*01
hs6M1-15											
EC2 81	Met ATG	Met ATG	Met ATG	Met ATG	Met ATG	Met ATG	Met ATG	Met ATG	Met ATG	Val GTG	Met ATG
TM7 279	Thr ACC	Thr ACC	Thr ACC	Thr ACC	Thr ACC	Thr ACT	Thr ACC	Thr ACC	Thr ACC	Thr ACT	Thr ACC
CP4 296	Asp GAC	Asp GAC	Asp GAC	Asp GAC	Asp GAC	Asn AAC	Asp GAC	Asp GAC	Asp GAC	Asp GAC	Asp GAC
	*01	*01	*01	*01	*01	*02	*01	*01	*01	*03	*01
hs6M1-16											
TM2 62	Ser TCC	Ser TCT	Ser TCT	Ser TCT	Ser TCT	Ser TCT	Ser TCT	Ser TCT	Ser TCT	Ser TCT	Ser TCT
TM2 63	Asn AAC	Asp GAC	Asp GAC	Asp GAC	Asp GAC	Asp GAC	Asp GAC	Asp GAC	Asp GAC	Asp GAC	Asp GAC
TM5 208	Pro CCT	Pro CCC	Pro CCT	Pro CCT	Pro CCT	Pro CCT	Pro CCT	Pro CCT	Pro CCT	Pro CCC	Pro CCT
	*02	*03	*01	*01	*01	*01	*01	*01	*01	*03	*01

(*Continues*)

Table 1. (*Continued*)

	BM28.7	BM19.7	LG2	KR3598	H2LCL	WT51	SA	YAR	OLGA	AMAI	PAC/BAC
hs6M1-17											
CP1 55	Gln CAG	STOP TAG	Gln CAG	Gln CAG	Gln CAG	Gln CAG	Gln CAG	Gln CAG	Gln CAG	Gln CAG	Gln CAG
TM2 61	Phe TTC	Phe TTT	Phe TTC	Phe TTC	Phe TTC	Phe TTC	Phe TTC	Phe TTC	Phe TTC	Phe TTC	Phe TTC
EC2 89	Arg CGC	Ser AGC	Arg CGC	Arg CGC	Arg CGC	Arg CGC	Arg CGC	Arg CGC	Arg CGC	Arg CGC	Arg CGC
CP2 121	Arg CGC	Cys TGC	Arg CGC	Arg CGC	Arg CGC	Arg CGC	Arg CGC	Arg CGC	Arg CGC	Arg CGC	Arg CGC
TM4 138	Arg CGG	Trp TGG	Arg CGG	Arg CGG	Arg CGG	Arg CGG	Arg CGG	Arg CGG	Arg CGG	Arg CGG	Arg CGG
TM4 160	Pro CCT	Ser TCT	Pro CCT	Pro CCT	Pro CCT	Pro CCT	Pro CCT	Pro CCT	Pro CCT	Ser TCT	Pro CCT
EC3 174	Pro CCG	Pro CCG	Pro CCG	Gln CAG	Pro CCG	Pro CCG	Pro CCG	Pro CCG	Gln CAG	Pro CCG	Pro CCG
TM6 246	Val GTG	Val GTG	Val GTG	Met ATG	Met ATG	Met ATG	Met ATG	Met ATG	Met ATG	Met ATG	Val GTG
TM6 254	Ala GCA	Ala GCC	Ala GCA	Ala GCA	Ala GCA	Ala GCA	Ala GCA	Ala GCA	Ala GCA	Ala GCA	Ala GCA
CP4 310	Met ATG	Arg AGG	Met ATG	Met ATG	Met ATG	Met ATG	Met ATG	Met ATG	Met ATG	Met ATG	Met ATG
	*01	*05	*01	*03	*02	*02	*02	*02	*03	*04	*01
hs6M1-18											
TM4 165	Ala GCC	Ala GCC	Ala GCC	Thr ACC	Ala GCC	Ala GCC	Ala GCC	Ala GCC	Ala GCC	Ala GCC	Ala GCC
	*01	*01	*01	*02	*01	*01	*01	*01	*01	*01	*01
hs6M1-19P											
EC3 186-190	del. 16 bp		del. 16 bp	del. - 16 bp	-	del. 16 bp	del. - 16 bp	-	del. 16 bp	del. 16 bp	del. 16 bp
	*01	*02	*01	*01 *02	*02	*01	*01 *02	*02	*01	*01	*01
hs6M1-20											
TM1 47	Val GTC	Phe TTC	Val GTC	Phe TTC	Phe TTC	Val GTC	Val GTC	Phe TTC	Val GTC	Val GTC	Val GTC
TM2 56	Leu CTT	Pro CCT	Leu CTT	Pro CCT	Pro CCT	Leu CTT	Leu Pro CTT CCT	Pro CCT	Leu Pro CTT CCT	Leu CTT	Leu CTT
TM3 104	Phe TTC	Phe TTC	Phe TTC	Phe TTC	Phe TTC	Ser TCC	Phe TTC	Phe TTC	Phe TTC	Phe TTC	Phe TTC
TM3 113	Phe TTC	Leu TTG	Phe TTC	Leu TTG	Leu TTG	Phe TTC	Phe Leu TTC TTG	Leu TTG	Phe Leu TTC TTG	Phe TTC	Phe TTC
CP2 120	Leu CTC	Arg CGC	Leu CTC	Arg CGC	Arg CGC	Leu CTC	Leu Arg CTC CGC	Arg CGC	Leu Arg CTC CGC	Leu CTC	Leu CTC
CP2 121	Ser TCT	Cys TGT	Ser TCT	Cys Ser TGT TCT	Cys TGT	Cys TGT	Ser Cys TCT TGT	Cys TGT	Ser Cys TCT TGT	Ser TCT	Ser TCT
TM4 159	Val GTA	Ile ATA	Val GTA	Ile ATA	Ile ATA	Val GTA	Ile Val ATA GTA	Ile ATA	Ile ATA	Val GTA	Val GTA
TM6 255	Leu CTT	Leu CTC	Leu CTT	Leu CTC	Leu CTC	Leu CTT	Leu CTC	Leu CTC	Leu CTC	Leu CTT	Leu CTT
	*01	*03	*01	*03$_a$ *04$_a$	*03	*02	*05$_a$ *06$_a$	*03	*05$_a$ *07$_a$	*01	*01
hs6M1-21											
EC1 23	Leu TTG	Leu TTG	Leu TTG	Leu TTG	Leu TTG	Leu TTG	Leu TTG	Leu TTG	Leu Trp TTG TGG	Leu TTG	Leu TTG
TM3 106	Phe TTC	Phe TTC	Phe TTC	Phe TTC	Phe TTC	Phe TTC	Phe TTC	Phe TTC	Phe TTC	Phe TTT	Phe TTC
CP3 233	Gly GGA	Gly GGA	Gly GGA	Gly GGA	Gly GGA	Gly GGA	Gly GGA	Gly GGA	Gly GGA	Arg AGA	Gly GGA
CP3 238	Phe TTT	Phe TTT	Phe TTT	Phe TTT	Phe TTT	Phe TTT	Phe TTC	Phe TTT	Phe TTT	Phe TTT	Phe TTT
	*01	*01	*01	*01	*01	*01	*02	*01	*01$_a$ *03$_a$	*04	*01
hs7M1-2											
TM4 137	His CAT	His CAT	His CAT	His Arg CAT CGT	His CAT	His CAT	His CAT	His Arg CAT CGT	His CAT	His Arg CAT CGT	His CAT
EC3 170	Arg AGA	Arg AGA	Arg Thr AGA ACA	Arg AGA	Arg AGA	Arg AGA	Arg AGA	Arg AGA	Arg AGA	Arg AGA	Arg AGA
	*01	*01	*01 *02	*01 *03	*01	*01	*01	*01 *03	*01	*01 *03	*01
hs17M1-20											
TM1 42	Ile Val ATT GTT	Ile Val ATT GTT	Ile ATT	Ile ATT	Ile ATT	Ile ATT	Ile ATT	Ile ATT	Ile Val ATT GTT	Ile ATT	Ile ATT
EC2 86	Ser Arg AGC AGG	Ser Arg AGC AGG	Ser AGC	Ser AGC	Ser AGC	Ser AGC	Ser AGC	Ser AGC	Ser Arg AGC AGG	Ser AGC	Ser AGC
TM3 120	Ile Val ATC GTC	Ile Val ATC GTC	Ile ATC	Ile ATC	Ile ATC	Ile ATC	Ile ATC	Ile ATC	Ile Val ATC GTC	Ile ATC	Ile ATC
EC3 168	Pro Thr CCT ACT	Pro Thr CCT ACT	Pro CCT	Pro CCT	Pro CCT	Pro CCT	Pro CCT	Pro CCT	Pro Thr CCT ACT	Pro Thr CCT ACT	Pro CCT
EC3 175	Asn Ser AAT AGT	Asn Ser AAT AGT	Asn AAT	Asn AAT	Asn AAT	Asn AAT	Asn AAT	Asn AAT	Asn Ser AAT AGT	Asn AAT	Asn AAT
EC3 193	Ala Ser GCC TCC	Ala Ser GCC TCC	Ala Ser GCC TCC	Ala Ser GCC TCC	Ser TCC	Ala Ser GCC TCC	Ser TCC	Ser TCC	Ala Ser GCC TCC	Ala Ser GCC TCC	Ala GCC
	*01$_a$ *03$_a$	*01$_a$ *03$_a$	*01 *02	*01 *02	*02	*01 *02	*02	*02	*01$_a$ *03$_a$	*01$_a$ *04$_a$	*01
hs19M1-4											
TM1 46	Ile ATC	Ile ATC	Ile ATC	Ile Thr ATC ACC	Ile Thr ATC ACC	Ile ATC	Ile ATC	Ile ATC	Ile Thr ATC ACC	Ile ATC	Ile ATC
	*01	*01	*01	*01 *02	*01 *02	*01	*01	*01	*01 *02	*01	*01

Nucleotide and amino acid substitutions and silent mutation within the different domains of the HLA-linked and non-HLA-linked OR genes leading to the indicated alleles. HLA class I haplotypes and ethnic origins of the analyzed cell lines are also indicated.
[a]Most likely allele.

OR Pseudogenes

Analysis of the PAC- or BAC-derived genomic sequences revealed that 17 of the HLA-linked OR loci, including several of them analyzed in detail on all 10 cell lines (*hs6M1-2P*, *-7P*, *-8P*, *-13P*, *-23P*, *-24P*, *-25P*, *-26P*) could be characterized as pseudogenes because of frameshifts, in-frame stop codons, or missing start codon. For several of these pseudogenes, between one and three alleles were found (data not shown), differing only by 1 or 2 nt in the respective cell lines (in the case of *hs6M1-5P*, *-24P*, *-25P*, and *-26P*, only the regions containing the stop codon or frameshifts were sequenced).

OR Genes With Open Reading Frames

Fourteen of the tested genes (*hs6M1-1*, *-3*, *-6*, *-10*, *-12*, *-15*, *-16*, *-17*, *-18*, *-20*, *-21*; *hs7M1-2*; *hs17M1-20*; *hs19M1-4*) showed an intact ORF-containing sequence for all domains as predicted by comparison with other known OR genes. Three of the HLA-linked genes appeared particularly interesting, as they exhibited intact ORFs in some of the cell lines but qualified as pseudogenes in others: *hs6M1-4P*, *-17*, and *-19P* (the gene status is defined here by the genomic sequence of the OR from the PAC or BAC used to obtain the original sequence information).

The 13 HLA-linked M-OR genes with ORF analyzed here (Table 1) exhibited 52 nucleotide substitutions altogether: 16 (31%) were silent and 36 (69%) resulted in a change in the amino acid sequence, while the *hs6M1-19P* gene exhibited a frameshift-generating deletion. Besides the *hs6M1-1* gene, which showed only a single silent mutation, all other genes exhibited at least two alleles with different amino acid sequences in the 10 cell lines analyzed. Apart from the *hs6M1-19P* locus, the amino acid replacements were always based on single nucleotide substitutions in the respective codons. The polymorphisms within the OR genes occurred at different positions (Table 1; Fig. 2): for example, *hs6M1-3* exhibited variations in TM3, CP3, and TM6, whereas in *hs6M1-6*, polymorphisms resulting in amino acid (AA) changes in TM2, 3, 4, and 5 were found, and *hs6M1-4P* was characterized by AA changes in the extracellular domains EC1 and EC3. However, exceptions to this finding were also observed: The variations of *hs6M1-3* (AA 113) and *hs6M1-6* (AA 111) in TM3 were located at corresponding AA positions and used the same codons, a situation similar to that of *hs6M1-20* (AA 104) and *hs6M1-21* (AA 106), although the latter gene exhibited only a silent mutation at this position.

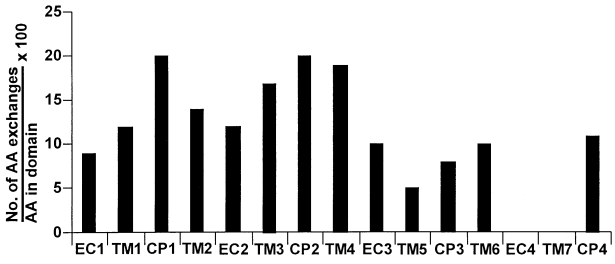

Figure 2 Frequency of polymorphisms leading to AA exchanges within all analyzed HLA-linked OR with ORF. The domains of the expressed proteins are plotted against the number of AA exchanges within these domains, expressed as a fraction of each domain's length.

Amino Acid Substitutions Within OR

The frequency of these AA substitutions varied considerably. For example, the Ala146Val substitution in TM4 of *hs6M1-6* was found on 11 of 19 chromosomes analyzed (including the PAC from which the sequence was derived), whereas other substitutions like the Ala111Thr change in TM3 of *hs6M1-6* were found only in the two haplotypes from the WT51 cell line (Table 1). We detected conservative substitutions where both amino acids of the OR have similar properties, as in the case of the Ala182Val exchange in EC3 of *hs6M1-4P* as well as drastic substitutions like the Gln234Arg exchange in CP3 of *hs6M1-10*, leading to a change in the polarity of the residue. The three OR on chromosomes 7, 17, and 19 exhibited very similar properties. Each of them revealed polymorphisms that resulted in both conservative as well as nonconservative AA changes in the expected proteins. AA substitutions were not found in each of the protein domains, but this could be because of the small number of non-HLA-linked genes analyzed so far.

OR Genes With Potentially Functional and Nonfunctional Alleles

In the case of the OR genes *hs6M1-4P* and *-17*, alleles (*hs6M1-4P* 01* and *-17* 05*, respectively) were found that are likely to give rise to nonfunctional protein products, while all other alleles at these loci are expected to be fully functional (Table 1). Yet another situation was observed for the *hs6M1-19P* gene: One of the two alleles found among the 10 cell lines and the original genomic sequence (*hs6M1-19P* 01*) had a 16-bp deletion within TM4, resulting in a frameshift and a presumably nonfunctional OR protein, while the other allele exhibited an intact ORF. In contrast to the *hs6M1-4P* and *-17* genes, no further polymorphisms were observed for this locus.

Numbers of OR Alleles

The number of alleles for each of the OR loci was also quite different (Table 2). While *hs6M1-1*, *-10*, *-18*, and *-19P* exhibited only two alleles (with the variation in *hs6M1-1* resulting in no amino acid exchange), the other OR genes had three (*hs6M1-6*, *-15*, and *-16*), four (hs6*M1-3*, *-12*, and *-21*), five (*hs6M1-4P* and *-17*), or even seven (*hs6M1-20*) alleles, thus approaching the

Table 2. OR Alleles and Haplotypes in the 10 Analyzed Cell Lines

Cell Line	OR gene – hs6M1-													HLA		OR Haplotype
	10	15	1	4P	3	6	21	20	19P	18	17	16	12	A	B	
H2LCL	*01	*01	*01	*01	*01	*01	*01	*03	*02	*01	*02	*01	*03	A3	B7	1
YAR	*01	*01	*01	*01	*01	*01	*01	*03	*02	*01	*02	*01	*03	A26	B38	1
LG2	*01	*01	*01	*01	*01	*01	*01	*01	*01	*01	*01	*01	*03	A2	B27	2
OLGA-1	*01	*01	*01	*01	*01	*01	*01	*05	*01	*01	*03	*01	*01	A31	B62	3
OLGA-2	*01	*01	*01	*02	*02	*02	*03	*07	*01	*01	*03	*01	*04	A31	B62	4
SA-1	*01	*01	*01	*01	*01	*01	*02	*05	*02	*01	*02	*01	*03	A24	B7	5
SA-2	*01	*01	*01	*01	*02	*02	*02	*06	*01	*01	*02	*01	*02	A24	B7	6
BM28.7	*01	*01	*01	*05	*02	*02	*01	*01	*01	*01	*01	*02	*01	A1	B35	7
BM19.7	*01	*01	*02	*05	*02	*02	*01	*03	*02	*01	*05	*03	*02	A2	B13	8
WT51-1	*01	*02	*01	*03	*03	*03	*01	*02	*01	*01	*02	*01	*01	A23	B65	9
WT51-2	*01	*02	*01	*03	*03	*03	*01	*02	*01	*01	*02	*01	*04	A23	B65	10
KR3598-1	*02	*01	*01	*01	*02	*01	*01	*03	*02	*02	*03	*01	*02	A2	B44	11
KR3598-2	*02	*01	*01	*02	*02	*02	*01	*04	*01	*02	*03	*01	*03	A2	B44	12
AMAI	*01	*03	*01	*04	*04	*02	*04	*01	*01	*01	*04	*03	*01	A68	B53	13
Number of alleles	2	3	2	5	4	3	4	7	2	2	5	3	4			

Note. In the case of the cell lines OLGA, SA, WT51, and KR3598, respectively, the two deducted haplotypes are indicated. However, only in the case of WT51 can an allele of a given OR gene be assigned to one of these haplotypes with certainty.

variability of the *HLA-A* gene in the cell line panel. The OR genes on chromosomes 7, 17, and 19 showed three, four, and two alleles and were therefore comparable with the HLA-linked OR loci. The allele definitions were not always unambiguous because of the presence in some of the cell lines of two alleles for a given locus. While this could be expected in our study for the non-HLA-linked genes, we had anticipated that the use of HLA-homozygous typing cells or hemizygous cell lines (BM19.7 and BM28.7 are reciprocal HLA haplotype loss mutants derived from the same maternal cell line) would circumvent this problem. Because of this complication, the allele designations are in some cases only tentative.

OR Haplotypes

The analysis given here allowed us also to define haplotypes of HLA-linked OR genes extending over a distance of nearly 2000 kb, from *hs6M1-10* to *hs6M1-12* (Table 2). A minimum of 13 different OR haplotypes could be deduced. Two cell lines (H2LCL and YAR) shared alleles for all HLA-linked OR loci investigated, although they exhibited distinct HLA class I haplotypes. The OR haplotypes of all other cell lines differed, and in the case of OLGA, SA, WT51, and KR3598, even the paternal and maternal haplotypes were distinguishable. However, certain combinations of neighboring alleles were conserved between several of the cell lines. For example, with the exception of the *hs6M1-20* and *-21* loci, one haplotype of SA was identical to that of H2LCL and YAR, and the BM19.7 and BM28.7 cell lines had the alleles of *hs6M1-4P*, *-3*, *-6*, and *-21* loci in common, while one of the OLGA and one of the KR3598 haplotypes, respectively, shared the alleles of *hs6M1-15*, *-1*, *-4P*, *-3*, and *-6*. The OR haplotypes of the cell lines WT51 and AMAI appeared to be the most divergent among the set of cell lines employed. However, even in these cases, the alleles of the closely linked OR loci *hs6M1-20* (only AMAI), *-19P*, and *-18* were observed at least partly also in LG2, OLGA (both haplotypes), SA (one haplotype), and BM28.7. The three different *HLA-A2* bearing haplotypes were associated with four different haplotypes of HLA-linked OR genes (Table 2).

DISCUSSION

Allelic variability of human M-OR genes has been predicted (Mombaerts 1999b) but not conclusively shown. Therefore, the demonstration of OR gene variability reported here is the first such systematic study and provides clear-cut evidence that sequence polymorphisms are a regular feature of these genes, irrespective of their chromosomal location. As HLA class I and II genes are characterized by extreme genetic polymorphism, with far more than 100 alleles for each of the HLA-A and -B genes, respectively (Bodmer et al. 1999), OR genes in close linkage to the HLA complex may be expected to exhibit pronounced variability as well, as demonstrated for the GABBR1 locus (Peters et al. 1998). Our finding that three of the HLA-linked OR genes analyzed exhibit alleles that are likely to be either functional or nonfunctinal, respectively, has also been foreseen (Mombaerts 1999b). The type of polymorphism observed by us is qualitatively different from that found by Trask and coworkers (1998). They described a 36-kb sequence containing three OR genes, one of them potentially functional, that may be in-

serted at various subtelomeric chromosomal sites, leading to a variable number of OR loci in individual genomes.

The major and minor M-OR gene clusters are separated from HLA-F by a distance of 135 kb and ~2000 kb, respectively (Fig. 1; for details, see Younger et al. 2000). Of the 34 M-OR genes within the two clusters, 17 (50%) are potentially functional at least in some of the haplotypes, while 50% are pseudogenes with unclear functional status. Another study (Rouquier et al. 1998b) arrived at an estimate of 72% pseudogenes among the M-OR loci within the human genome, while the OR gene cluster in the human chromosomal region 17p13.3 contained nearly double as many genes with ORF than with pseudogenes (Glusman et al. 2000). The frequency of potentially functional genes is greater in the major than in the minor HLA-linked cluster (14/25 [56%] vs. 3/9 [33%]). However, these numbers are likely to be underestimates. First, inspection of further alleles of the pseudogenes may yield variants with ORF; second, the search of EST databases has revealed that at least the *hs6M1-14P* and *-24P* genes are transcribed (Younger et al. 2000). Finally, the expression of functional chemokine receptors despite loss of the first two transmembrane domains (Ling et al. 1999) may indicate that OR genes lacking start codons (e.g., *hs6M1-14P*) could be expressed also at the protein level by using an alternative ATG start codon at the beginning of extracellular domain 2. In this context, the finding of two ESTs splicing to this methionine in the *hs6M1-16* gene is remarkable, resulting in a possibly functional, N-terminally truncated OR gene product (Younger et al. 2000). Similarly, it could be that the *hs6M1-17* 05* allele, which exhibits a stop codon in the first cytoplasmic domain (Table 1), is in fact functional.

However, this situation is likely to be different for the *hs6M1-4P*01* and *-19P*01* alleles (Table 1). In the first case, the stop codon occurs within the third extracellular domain, while *hs6M1-19P*01* exhibits a sizable deletion, also within this domain. In either case, the resulting C-terminally truncated protein is highly likely to be nonfunctional. The frequency of the two nonfunctional alleles, respectively, is quite high among the cell lines analyzed: In both cases, more than half of the chromosome 6s carried these sequence polymorphisms. This is in contrast to the *hs6M1-17*05* allele, which occurred only in the HLA hemizygous BM19.7 cell line. The *hs6M1-19P*01* allele must have been derived from the potentially functional *hs6M1-19P*02* allele, as it appears very unlikely that an insertion of 16 bp, at the correct position, into the **01* allele would have occurred to restore the **02* allele. Therefore, the *hs6M1-19P* gene is an example of pseudogenization within the human species, not just between species as described by Rouquier and colleagues (1998a). Clearly, more work is needed to evaluate the frequency with which comparable events occur in further individuals and other OR genes.

The allelic differences observed for nearly all of the OR genes studied here could affect their interaction with extracellular ligands or intracellular proteins of the signal transduction cascade. By comparing sequence features of 197 OR from various species, Pilpel and Lancet (1999) have found that the regions of highest sequence diversity among these OR are located within the transmembrane regions 3, 4, 5, and 6, although the extracellular regions 2 and, in particular, 4 (EC1 and EC3 in their nomenclature) are also highly variable. However, in the set of HLA-linked OR studied here, EC4 exhibits no variability at all (Fig. 2), and only a silent polymorphism is observed within the TM7 region in *hs6M1-15* (Table 1). The exchanges in the transmembrane regions could have functional consequences, as it has been suggested that these domains are involved in ligand binding (Buck and Axel 1991; Pilpel and Lancet 1999). Even conservative AA exchanges like the Val159Ile polymorphism in TM4 of *hs6M1-20* might well give rise to changes in ligand specificity, as this has been shown for a similar AA exchange (Val206Ile, in TM5 of a murine OR), which leads to preferential binding of either octanal or heptanal (Krautwurst et al. 1998). The Arg138Trp polymorphism in the *hs6M1-17* gene can serve as an example for several of the nonconservative exchanges that are even more likely to result in a functional difference. Furthermore, M-OR genes contain intracellular AA-sequences that are highly conserved and are thought to be important for receptor function, possibly G-protein binding (Hedin et al. 1993). This is exemplified by the ORIC1, ORIC2, and ORIC3 sequence motifs in cytoplasmic loops CP1, 2, and 3 (Pilpel and Lancet 1999). The *hs6M1-17* gene exhibits changes in the first two of these motifs (in CP1, a Gln55His replacement in nearly all alleles, and in CP2, an Arg121Cys polymorphism), while the *hs6M1-10* and *hs6M1-21* genes were polymorphic in CP3 (Table 1).

The task of reliably distinguishing thousands of odors might best be accomplished by OR that are not only numerous but also polymorphic. However, in the absence of any data on ligand specificity for the HLA-linked OR, it is clearly premature to embark on speculations regarding the possible importance of AA replacements within any of the OR domains. Only those OR with potentially functional and nonfunctional alleles like *hs6M1-4P* allow us to predict that individuals who are homozygous for the *hs6M1-4P*01* allele may exhibit a specific anosmia. It is evident that this olfactory deficit offers interesting opportunities to correlate behavioral features with genetic polymorphisms (see also Mombaerts 1999b). Persons heterozygous for an active and an inactive allele might exhibit altered

odorant detection thresholds than people homozygous for the active allele. It is, however, possible that the situation will not be as simple, as the specificities of different OR proteins might overlap, so that an inactive OR could be compensated for by related receptors, possibly even within the same OR cluster (Malnic et al. 1999; Tsuboi et al. 1999).

For the first time, the data presented here allow also the definition of haplotypes for any of the known OR gene clusters (Table 2). Despite limited OR polymorphism, when compared to HLA class I loci, it was possible to demonstrate the presence of 13 distinct haplotypes of HLA-linked OR genes among the cell lines analyzed, which represent 18 different chromosome 6s. As pointed out before, the number of haplotypes is higher than that of the HLA class I haplotypes within the cell line panel. Despite the consanguineous origin of most of the cell lines employed here, it can not be excluded that some are in fact not HLA class I homozygous, possibly leading to an increase in OR gene polymorphism, as well. Obviously, candidates for this constellation are OLGA, SA, WT51, and KR3598, which are characterized by one (WT51) to six (OLGA) HLA-linked OR loci with two alleles. These numbers are subject to further changes when all HLA-linked OR genes will have been discovered and additional polymorphisms detected. The analysis of the OR gene haplotypes reveals also that the cell lines H2LCL and YAR share identical OR alleles for all loci analyzed, with the sole exception of the *hs19M1-4* gene, where H2LCL but not YAR exhibits heterozygosity. YAR is of Jewish origin, while we could not trace the origin of H2LCL (described to be of Caucasian origin) with certainty.

It is currently unclear whether the sharing of blocks of alleles by different HLA-linked OR haplotypes (Table 2), which to some extent reminds us of the organization of the HLA class I region into three genomic blocks (reviewed by Kulski et al. 2000), will be observed also after further individuals have been analyzed. The CEPH families provide a good opportunity for such studies: They have been HLA typed, and extended haplotypes caused by linkage disequilibrium are known to exist between the *HLA-A* and *HFE* genes in these families (Malfroy et al. 1997). As MHC class I and/or linked loci are involved in shaping individual-specific odors and odor preferences (Wedekind et al. 1995; Wedekind and Füri 1997; Penn and Potts 1998b; Milinski and Wedekind 2000), further research in this area might also seek to provide evidence in favor of or against a functional connection between linked HLA and OR genes.

METHODS

Cell Culture

Cell lines were derived from different donors, representing different HLA haplotypes and different ethnic origins. Eight of the 10 cell lines were HLA homozygous, whereas two (BM19.7, BM28.7) were HLA hemizygous (Ziegler et al. 1985; Volz et al. 1992). All cell lines were grown in RPMI 1640 medium containing antibiotics and 10% fetal calf serum.

Polymerase Chain Reaction

We designed two to three pairs of primers for the respective genes (*hs6M1-1* = AL022727; *hs6M1-2P* = AL022727, AJ132194; *hs6M1-3* = AL022727; *hs6M1-4P* = AL022727; *hs6M1-5P* = AL022727; *hs6M1-6* = AL022727; *hs6M1-7P* = AL022727; *hs6M1-8P* = CAB55431; *hs6M1-10* = Z98744; *hs6M1-12* = AL031983, AC006137; *hs6M1-13P* = AL031983, AC006137; *hs6M1-15* = AL035402; *hs6M1-16* = AL035542, AC004178; *hs6M1-17* = AL035542; *hs6M1-18* = AL035542; *hs6M1-19P* = AL035542; *hs6M1-20* = AL035542; *hs6M1-21* = AL096770; *hs6M1-23P* = AL050339; *hs6M1-24P* = AL050339; *hs6M1-25P* = AL050339; *hs6M1-26P* = AL035402; *hs7M1-1* = AC004853; *hs17M1-20* = AC002085; and *hs19M1-4* = AC002988), resulting in overlapping PCR products. Because two different sequencing strategies were employed, PCR primers were generated with or without M13 tail. Specificity of the primers was tested by aligning all known ORs.

PCR primers were generated with M13 for/rev tails to sequence the PCR products with fluorescence-labeled M13 primers. For PCR, ~100 ng template DNA, 10 pmole of each primer, 0.2 mM of each dNTP (Pharmacia Biotech), 1 U Ampli-Taq DNA-polymerase (Perkin-Elmer), and 1× buffer (Perkin Elmer) were used in a final volume of 20 µL.

PCR primers without M13 tail were used for subsequent sequencing with the ABI Prism cycle sequencing kit. For PCR, ~50 ng template DNA, 100 ng of each primer, 2 mM of each dNTP (GIBCO), 1–2 mM $MgCl_2$, 1 U Taq polymerase (GIBCO) and 1× KCl buffer was used in a final volume of 50 µL. The primers and PCR conditions are given in Table 3.

Cycle Sequencing

We used two different sequencing strategies. Sequence analysis was performed either using the Thermo Sequenase fluorescent labeled primer cycle sequencing kit with 7-deaza-dGTP (Amersham Pharmacia Biotech) employing 150 ng DNA for 1 kb, 1 pmol primer (M13for: 5′-TGTAAAACGACGGCCAGT; M13rev: 5′-CAGGAAACAGCTATGACC), 0.25 µL DMSO and 2 µL reaction mix in a final volume of 7 µL (27 cycles; 95°C for 4 min, 95°C for 15 sec, 56°C for 15 sec, 70°C for 15 sec) followed by analysis on 4% polyacrylamide gels with a LI COR sequencer (MWG Biotech) or by employing 3.2 pmol of each of the PCR primers using the ABI Prism cycle sequencing kit. In this case, the DNA products were electrophoresed through 6% polyacrylamide gels in an ABI semiautomatic sequencer.

Restriction Analysis

All nucleotide substitutions observed by ABI cycle sequencing were subsequently confirmed by restriction analysis. Sequences of the two alternate alleles were restriction mapped using the tacg (v2.38) program at University of California Irvine. Restriction enzymes detecting the polymorphic site were selected and used to digest 10 µL of the PCR products generated from the panel of DNAs in a 50-µL reaction overnight. Digestion products were then analyzed on a 2.5% agarose gel.

Cloning

Most of the genes showing heterozygous positions in different

Table 3. PCR Primers and PCR Conditions

1.Primers with M13 tail and PCR-parameters

Gene-Name	Primer-Name	Sequence	Temp °C	DMSO
hs6M1-1	OR1-4for	*TGTAAAACGACGGCCAGT*ATAAACAAACATTGATTGCT	50	+2%
	OR1-615rev	*CAGGAAACAGCTATGACC*CATGACAACTTGAGAAGTGC		
hs6M1-1	OR1-477for	*TGTAAAACGACGGCCAGT*AATTATTGGTTCTGCCTAAG	57	+2%
	OR1-1064rev	*CAGGAAACAGCTATGACC*TGAGTTCAAAGGTCATTACC		
hs6M1-2P	OR2-77F	*TGTAAAACGACGGCCAGT*TTCTGACCCCAGGTTACTGC	60	
	OR2-764R	*CAGGAAACAGCTATGACC*GAACCCTTTTCACCACAGGC		
hs6M1-2P	OR2-683F	*TGTAAAACGACGGCCAGT*ACTGCCACATTACAATTGCC	55	
	OR2-1231R	*CAGGAAACAGCTATGACC*TTGCAAGAACATGTAAAGCG		
hs6M1-3	OR3-27for	*TGTAAAACGACGGCCAGT*GTGTGCTGATATTTTTGGAT	54	
	OR3-711rev	*CAGGAAACAGCTATGACC*AATATGGAGCTTGTGATCAT		
hs6M1-3	OR3-576for	*TGTAAAACGACGGCCAGT*AACTCAGCACTTCATTCCTC	64	
	OR3-1127rev	*CAGGAAACAGCTATGACC*GTGCTAACAGAGTGCGTGAA		
hs6M1-4P	OR4-446f	*TGTAAAACGACGGCCAGT* TACAGTTCAACTTTACTTTG	56	
	OR4-1274r	*CAGGAAACAGCTATGACC*AGAACAAAATGGTACTAATC		
hs6M1-4P	OR4-48	*TGTAAAACGACGGCCAGT*ATTGGGATACTTTTTTCTCC	61	
	OR4-653	*CAGGAAACAGCTATGACC*GGCGATGTCTACATAGGGGT		
hs6M1-5P	OR5-419f	*TGTAAAACGACGGCCAGT*GGTCAGTCTCTGGGGTGTGG	60	
	OR5-1231r	*CAGGAAACAGCTATGACC*GTTACCAGGATCTCCACGAC		
hs6M1-6	OR6-47F	*TGTAAAACGACGGCCAGT*AAGTGAGCGGTTGACAATGC	64	+2%
	OR6-699R	*CAGGAAACAGCTATGACC*GGTCAGCTCATTTGCATGGG		
hs6M1-6	OR6-621F	*TGTAAAACGACGGCCAGT*TACCCCTTTGTGGACATCGC	60	+2%
	OR6-1112Rn	*CAGGAAACAGCTATGACC*GGAAACCACCTTTCAAGATG		
hs6M1-7P	OR7-85F	*TGTAAAACGACGGCCAGT*TCATATACCACCCGTCTTCC	60	+2%
	OR7-726R	*CAGGAAACAGCTATGACC*GAGAGCTGAATCAGAGCTGG		
hs6M1-7P	OR7-643F	*TGTAAAACGACGGCCAGT*CCCAGTCTCTGATCCAGTCC	67	
	OR7-1280R	*CAGGAAACAGCTATGACC*TATGAAGAATTTAAGGGCCC		
hs6M1-10	OL-AE 800 f	*TGTAAAACGACGGCCAGT*TTTTTCACATTATTAATTGG	52	
	OL-AE 1525 r	*CAGGAAACAGCTATGACC*ATGAAGAATAGTTCAGCCTC		
hs6M1-10	OL-AE 1446 f	*TGTAAAACGACGGCCAGT*TGGTCACAAAGAAGTGGATC	52	
	Ol-AE 2164 r	*CAGGAAACAGCTATGACC*TGATTTCTGGATCAGAAAGG		
hs6-M1-15	OR15-403f	*TGTAAAACGACGGCCAGT*TCTCCCTCCTCCTTCTCCTC	60	
	OR15-1029r	*CAGGAAACAGCTATGACC*ATCCAGAATGTTGTTTCCAC		
hs6-M1-15	OR15-847f	*TGTAAAACGACGGCCAGT*CTCAGTTGAGTGCCTTCTCC	60	
	OR15-1708r	*CAGGAAACAGCTATGACC*ATTTGTGTTTTATTTCAAGC		
hs6M1-16	OR11-204f	*TGTAAAACGACGGCCAGT*GATCAGAAGGAACAGGGAAC	58	
	OR11-1091r	*CAGGAAACAGCTATGACC*CCCAAAGGCCTTTCTCCATG		
hs6M1-16	Fat11 1038	*TGTAAAACGACGGCCAGT*CAGCTCTAATTCGACTCTCC	52	
	Fat11 B	*CAGGAAACAGCTATGACC*GAAATCTATAGGAGTGATGA		
hs6M1-18	OR18-108-f	*TGTAAAACGACGGCCAGT*GTTGAGGAAGGAATGACAAC	60	
	OR18-756-r	*CAGGAAACAGCTATGACC*CATCTACCACAAATCCAGAG		
hs6M1-18	OR18-606f	*TGTAAAACGACGGCCAGT*GTTCTTTATCTTCGGCTCTC	53	
	OR18-1368r	*CAGGAAACAGCTATGACC*TTTCATTTTTAGTATAACTG		
hs6M1-24P	OR24-f713	*TGTAAAACGACGGCCAGT*GTCAAAATTGCATGTGGGGC	60	
	OR24-r1663	*CAGGAAACAGCTATGACC*ACAGTGCTGGGATTACAGGC		
hs6M1-25P	OR25-f222	*TGTAAAACGACGGCCAGT*GCCCCACCTGGAAAAGATCC	62	
	OR25-r866	*CAGGAAACAGCTATGACC*CCCACATGTCCCCAGGCCTTT		
hs6M1-26P	OR26-f53	*TGTAAAACGACGGCCAGT*CCATCTGCAAACACTTGAGG	60	
	OR26-r573	*CAGGAAACAGCTATGACC*ATACAAAAAGACAAGAAGCC		

The standard cycling conditions are: 94°C 5min, 94°C 30 sec, annealing temperature x°C 30 sec, 72°C 40 sec, 72°C 7

2. Primers without M13 tail and PCR-parameters:

Gene-Name	Primer-Name	Sequence	Temp °C	$MgCl_2$ [mM]
hs6M1-8P	OLFR8pF	GTTGGCTGTGATGGCCTATG	62	1
	OR8pR	GATGGGAAGGTTAAGGCTGG		
hs6M1-12	OLFR12F1	TTACTACATTTCAGTCGCTGTC	62	1,5
	OLFR12R1	GTGCTGAGGATTAACTCTGC		
hs6M1-12	OLFR12F2	GTGACCACAGTGAGATGGG	60	1
	OLFR12R2	ATCAGCTTCCTGGACTGCTC		
hs6M1-12	OLFR12F3	CTGTCAAGAGGATGCACTCAG	62	1,5
	OLFR12R3	AGACAGGTTGAATCACACTGG		
hs6M1-13P	OLFR13F1	CTTGGGAGCTCAAACTTGTTC	62	1,5
	OLFR13R1	CACCTCCTACAATGAGATCCAG		
hs6M1-13P	OLFR13F2	GAGGCACAGCCAAGATGAAG	62	1
	OLFR13R2	GGACCTCTGTTTCACCATAAGTTG		
hs6M1-13P	OLFR13F3	GCAGCCCAGTAAGATGATGG	62	1
	OLFR13R3	AACACTCACCTACTGGGACCTC		
hs6M1-17	OR17-F	TTGTCTTTCTGACAGGCTGG	60	2
	OR17-R	AGGGAGATCTAGTGCTGCGA		
hs6M1-19P	OLFR19pF	ATGAAGTGGGAGGCACAAGT	62	1,5
	OLFR19pR	GCTGCACTCCCTAATGACCT		
hs6M1-19P	OLFR19.1F	TTTTATCCAGTTCCCTCTGTTG	60	1,5
	OLFR19.1R	ATTCTTTTAGCATGCTCCGC		
hs6M1-19P	OLFR19.2F	AGCAATGGCTTCACATCACAG	60	1
	OLFR19.2R	CCTGGACATCTGCTACTCCA		
hs6M1-19P	OLFR19.3F	GGCTTATGCATCCCAAGAA	60	1,5
	OLFR19.3R	TCTGAGTCGGAAGGATTCTGA		
hs6M1-20	OR20-F	TCCCCAGAAGAAAGAAATACGT	60	1
	Or20-R	GGCTAGTGTCTTGCATTTTCAA		
hs6M1-21	OLFR21.1F	TTCTCTTTGCCCAATTCCTG	60	1
	OLFR21.1R	TATGCCATTGCTGCCAGTAG		
hs6M1-21	OLFR21.2F	GGGGTGTGTGGTTCAACTTT	62	1
	OLFR21.2R	CGTCTTCCCTCTGAGGACTG		
hs6M1-21	OLFR21.3F	CATTGGTTGGACTCCTTTCC	60	1,5
	OLFR21.3R	TCAATATCTGTCCCAACATTGC		
hs6M1-23P	OLFR23pF	CTTGTGGTCATCTCCTGGGT	60	1,5
	OLFR23pR	AGCCACAGCAATGAAACCAT		
hs7M1-2	OLFR 7F1	TAGGGCTGGCTGGCATGTA	65	1,5
	OLFR 7R1	ACAGCTCTGGAATGGGATGG		
hs7M1-2	OLFR7F3	GTGGACACCTCCTCCAATGA	64	1,5
	OLFR 7R3	CTGGGTGGAGAACACTGAGG		
hs7M1-2	OLFR 7F1	TAGGGCTGGCTGGCATGTA	62	1,5
	OLFR 7R3	CTGGGTGGAGAACACTGAGG		
hs17M1-20	OLFR 17F1	ACACCTCATCCTGCTTCTGC	64	1,5
	OLFR 17R1	ATGGCCAGATAGCGGTCATA		
hs17M1-20	OLFR 17F2	CCTATAAGGCCTGCCTCTCC	64	1,5
	OLFR 17R2	CATAGAAGATGCCCACCACA		
hs17M1-20	OLFR 17F3	ACTGTGTCCTATGCCCATGT	62	1,5
	OLFR 17R3	GTGCTGTTGGTGAGAAGCTG		
hs19M1-4	OLFR 19F1	GGCACAGAGTGAGAGACCC	58	1
	OLFR 19R1	GATCATAGGCCATCACAGCC		
hs19M1-4	OLFR 19F2	GAGTCATCACCTATGCAGGCT	60	1
	OLFR 19R2	TGCCTTGTACTTCCCCTGAG		
hs19M1-4	OLFR 19F3	TGGGGATCCTTTGCTCTTAC	58	2
	OLFR 19R3	GCAAATTCCACTTTCACAACC		

The standard cycling conditions are: +x mM $MgCl_2$, 34 cycles; 96°C 2min, 92°C 30 sec,

PCR fragments after sequencing with fluorescence-labeled M13 primers were amplified by PCR with both exterior primers to a DNA fragment encoding the full-length receptor. The fragments were cloned into vector pCR II-TOPO (Invitrogen) according to the manufacturer's recommendations and grown overnight in 200 µL LB medium with ampicillin. The full-length insert was recovered by PCR with the same primers and sequenced to determine the respective alleles of the genes.

Nomenclature

In the absence of an official OR nomenclature, we have as-

signed unique designations to describe OR genes unambiguously (Younger et al. 2000; Ziegler et al. 2000a). The names identify the species (hs for humans [*Homo sapiens*], mm for mouse [*Mus musculus*]), followed by a number representing the chromosome (e.g. "6" for chromosome 6), then a letter and a number indicating the OR type and subtype, ("M1" for MOE subtype 1; "V1" and "V2" for the two VNO subtypes). This descriptive information is followed by a dash (−) and an arbitrary but unique gene identification number. Pseudogenes are indicated with a "P" (e.g., *hs6M1-4P*). Following the HLA nomenclature, OR alleles are indicated by an asterisk (*) and a unique allele number (*hs6M1-20*07*). The allele **01* was used here to indicate the allele in the originally sequenced genomic DNA. Furthermore, OR genes and proteins may be referred to as "M-OR", "V1-OR," or "V2-OR" depending on the type and subtype they belong to. This proposal for a consistent OR gene nomenclature was discussed and submitted to the HUGO/GDB Nomenclature Committee for consideration.

ACKNOWLEDGMENTS

We thank the Volkswagen Foundation (grant I/72 740 to A.Z. and J.T.) and the Wellcome Trust (S.B.) for financial support. R.M.Y. was supported by a studentship of the UK Medical Research Council. A.Z. and S.B. thank the Wellcome Trust for a Biomedical Research Collaboration Grant.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Beauchamp, G.K. and Yamazaki, K. 1997. HLA and mate selection in humans: Commentary. *Am. J. Hum. Genet.* **61:** 494–496.

Bodmer, J.G., Marsh, S.G., Albert, E.D., Bodmer, W.F., Bontrop, R.E., Dupont, B., Erlich, H.A., Hansen, J.A., Mach, B., Mayr, W.R., et al. 1999. Nomenclature for factors of the HLA system, 1998. *Vox Sang* **77:** 164–191.

Brown, R.E., Roser, B., and Singh, P.B. 1989. Class I and class II regions of the major histocompatibility complex both contribute to individual odors in congenic inbred strains of rats. *Behav. Genet.* **19:** 659–674.

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

Dulac, C. and Axel, R. 1995. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83:** 195–206.

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.

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

Gruen, J.R., Nalabolu, S.R., Chu, T.W., Bowlus, C., Fan, W.F., Goei, V.L., Wei, H., Sivakamasundari, R., Liu, Y., Xu, H.X., et al. 1996. A transcription map of the major histocompatibility complex (MHC) class I region. *Genomics* **36:** 70–85.

Hedin, K.E., Duerson, K., and Clapham, D.E. 1993. Specificity of receptor-G protein interactions: searching for the structure behind the signal. *Cell Signal* **5:** 505–518.

Herrada, G. and Dulac, C. 1997. A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. *Cell* **90:** 763–773.

Krautwurst, D., Yau, K.W., and Reed, R.R. 1998. Identification of ligands for olfactory receptors by functional expression of a receptor library. *Cell* **95:** 917–926.

Kulski, J.K., Gaudieri, S., and Dawkins, R.L. 2000. Transposable elements and the metamerismatic evolution of the HLA class I region. In *Major histocompatibility complex—Evolution, structure, function* (ed. M. Kasahara), pp. 158–177. Springer, Tokyo.

Leinders-Zufall, T., Lane, A.P., Puche, A.C., Ma, W., Novotny, M.V., Shipley, M.T., and Zufall, F. 2000. Ultrasensiive pheromone detection by mammalian vomeronasal neurons. *Nature* **405:** 792–796.

Ling, K., Wang, P., Zhao, J., Wu, Y.-L., Cheng, Z.-J., Wu, G.-X., Hu, W., Ma, L., and Pei, G. 1999. Five-transmembrane domains appear sufficient for a G protein-coupled receptor: Functional five-transmembrane domain chemokine receptors. *Proc. Natl. Acad. Sci.* **96:** 7922–7927.

Malfroy, L., Roth, M.P., Carrington, M., Borot, N., Volz, A., Ziegler, A., and Coppin, H. 1997. Heterogeneity in rates of recombination in the 6-Mb region telomeric to the human major histocompatibility complex. *Genomics* **43:** 226–231.

Malnic, B., Hirono, J., Sato, T., and Buck, L.B. 1999. Combinatorial receptor codes for odors. *Cell* **96:** 713–723.

Matsunami, H. and Buck, L.B. 1997. A multigene family encoding a diverse array of putative pheromone receptors in mammals. *Cell* **90:** 775–784.

Milinski, M. and Wedekind, C. 2000. Evidence for MHC-correlated perfume preferences in humans. *Behav. Ecol.*, in press.

Mombaerts, P. 1999a. Molecular biology of odorant receptors in vertebrates. *Annu. Rev. Neurosci.* **22:** 487–509.

———. 1999b. Odorant receptor genes in humans. *Curr. Opin. Genet. Dev.* **9:** 315–320.

Naruse, T.K., Nose, Y., Ando, R., Araki, N., Shigenari, A., Ando, A., Ishihara, M., Kagiya, M., Nabeya, N., Isshiki, G., et al. 1998. Extended HLA haplotypes in Japanese homozygous typing cells. *Tissue Antigens* **51:** 305–308.

O'Shea, S.F., Chaure, P.T., Halsall, J.R., Olesnicky, N.S., Leibbrandt, A., Connerton, I.F., and Casselton, L.A. 1998. A large pheromone and receptor gene complex determines multiple B mating type specificities in *Coprinus cinereus*. *Genetics* **148:** 1081–1090.

Penn, D. and Potts, W.K. 1998a. Untrained mice discriminate MHC-determined odors. *Physiol. Behav.* **64:** 235–243.

———. 1998b. How do major histocompatibility complex genes influence odor and mating preferences? *Adv. Immunol.* **69:** 411–436.

———. 1998c. MHC-disassortative mating preferences reversed by cross-fostering. *Proc. R. Soc. Lond. B Biol. Sci.* **265:** 1299–1306.

Peters, H.C., Kammer, G., Volz, A., Kaupmann, K., Ziegler, A., Bettler, B., Epplen, J.T., Sander, T., and Riess, O. 1998. Mapping, genomic structure, and polymorphisms of the human GABABR1 receptor gene: Evaluation of its involvement in idiopathic generalized epilepsy. *Neurogenetics* **2:** 47–54.

Pilpel, Y. and Lancet, D. 1999. The variable and conserved interfaces of modeled olfactory receptor proteins. *Protein Sci.* **8:** 969–977.

Potts, W.K. and Wakeland, E.K. 1993. Evolution of MHC genetic diversity: A tale of incest, pestilence and sexual preference. *Trends Genet.* **9:** 408–412.

Potts, W.K., Manning, C.J., and Wakeland, E.K. 1991. Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature* **352:** 619–621.

Rouquier, S., Friedman, C., Delettre, C., van den Engh, G., Blancher, A., Crouau-Roy, B., Trask, B.J., and Giorgi, D. 1998a. A gene recently inactivated in human defines a new olfactory receptor family in mammals. *Hum. Mol. Genet.* **7:** 1337–1345.

Rouquier, S., Taviaux, S., Trask, B.J., Brand-Arpon, V., van den Engh, G., Demaille, J., and Giorgi, D. 1998b. Distribution of olfactory receptor genes in the human genome. *Nat. Genet.* **18:** 243–250.

Singh, P.B., Brown, R.E., and Roser, B. 1987. MHC antigens in urine as olfactory recognition cues. *Nature* **327:** 161–164.

Tay, G.K., Cattley, S.K., Chorney, M.J., Hollingsworth, P.N., Roth, M.P., Dawkins, R.L., and Witt, C.S. 1997 Conservation of ancestral haplotypes telomeric of HLA-A. *Eur. J. Immunogenet.* **24:** 275–285.

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., et al. 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.

Tsuboi, A., Yoshihara, S.-I., Yamazaki, N., Kasai, H., Asai-Tsuboi, H., Komatsu, M., Serizawa, S., Ishii, T., Matsuda, Y., Nagawa, F. et al. 1999. Olfactory neurons expressing closely linked and homologous odorant receptor genes tend to project their axons to neighbouring glomeruli on the olfactory bulb. *J. Neurosci.* **19:** 8409–8418.

Volz, A., Fonatsch, C., and Ziegler, A. 1992. Regional mapping of the gene for autosomal dominant spinocerebellar ataxia (SCA1) by localizing the closely linked D6S89 locus to 6p24.2-p23.05. *Cytogenet. Cell Genet.* **60:** 37–39.

Wedekind, C. and Füri, S. 1997. Body odor preferences in men and women: Do they aim for specific MHC combinations or simply heterozygosity? *Proc. R. Soc. Lond. B Biol. Sci.* **264:** 1471–1479.

Wedekind, C., Seebeck, T., Bettens, F., and Paepke, A.J. 1995. MHC-dependent mate preferences in humans. *Proc. R. Soc. Lond. B Biol. Sci.* **260:** 245–249.

Wysocki, C.J. 1989. Vomeronasal chemorecognition: Its role in reproductive fitness and physiology. In *Neural Control of Reproductive Function*, pp. 545–566, Liss, New York.

Yamazaki, K., Boyse, E.A., Mike, V., Thaler, H.T., Mathieson, B.J., Abbott, J., Boyse, J., Zayas, Z.A., and Thomas, L. 1976. Control of mating preferences in mice by genes in the major histocompatibility complex. *J. Exp. Med.* **144:** 1324–1335.

Yamazaki, K., Yamaguchi, M., Baranoski, L., Bard, J., Boyse, E.A., and Thomas, L. 1979. Recognition among mice: Evidence from the use of a Y-maze differentially scented by congenic mice of different major histocompatibility types. *J. Exp. Med.* **150:** 755–760.

Yamazaki, K., Beauchamp, G.K., Matsuzaki, O., Kupniewski, D., Bard, J., Thomas, L., and Boyse, E.A. 1986. Influence of a genetic difference confined to mutation of H-2K on the incidence of pregnancy block in mice. *Proc. Natl. Acad. Sci.* **83:** 740–741.

Yamazaki, K., Beauchamp, G.K., Singer, A., Bard, J., and Boyse, E.A. 1999. Odortypes: Their origin and composition. *Proc. Natl. Acad. Sci.* **96:** 1522–1525.

Zhao, H., L. Ivic, J.M. Otaki, M. Hashimoto, K. Mikoshiba, and S. Firestein. 1998. Functional expression of a mammalian odorant receptor. *Science* **279:** 237–242.

Ziegler, A., Ehlers, A., Forbes, S., Trowsdale, J., Uchanska-Ziegler, B., Volz, A., Younger, R., and Beck, S. 2000a. Polymorphic olfactory receptor genes and HLA loci constitute extended haplotypes. In *Major Histocompatibility Complex–evolution, Structure, and Function*. (M. Kasahara, ed.) pp. 110–130. Springer Verlag Tokyo.

Ziegler, A., Ehlers, A., Forbes, S.A., Trowsdale, J., Volz, A., Younger, R., and Beck, S., 2000b. Polymorphism in olfactory receptor genes: A cautionary note. *Hum. Immunol.* (in press).

Ziegler, A., Müller, C., Heinig, J., Radka, S.F., Kömpf, J., and Fonatsch, C. 1985. Monosomy 6 in a human lymphoma line induced by selection with a monolonal antibody. *Immunobiology* **169:** 455–460.

Received August 4, 2000; accepted in revised form October 12, 2000.