Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 346 (2005) 13–21
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

Comparative evolutionary analysis of olfactory receptor gene clusters between humans and mice

Yoshihito Niimura*, Masatoshi Nei

Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University, 328 Mueller Laboratory, University Park, PA 16802, USA

Received 1 July 2004; received in revised form 17 September 2004; accepted 28 September 2004
Available online 7 January 2005
Received by T. Gojobori

Abstract

Olfactory receptor (OR) genes form the largest multigene family in mammalian genomes. Humans have ~800 OR genes, but >50% of them are pseudogenes. By contrast, mice have ~1400 OR genes and pseudogenes are ~25%. To understand the evolutionary processes that shaped the difference of OR gene families between humans and mice, we studied the genomic locations of all human and mouse OR genes and conducted a detailed phylogenetic analysis using functional genes and pseudogenes. We identified 40 phylogenetic clades with high bootstrap supports, most of which contain both human and mouse genes. Interestingly, a particular clade contains ~100 pseudogenes in humans, whereas the numbers of pseudogenes are <20 for most of the mouse clades. We also found that the organization of OR genomic clusters is well conserved between humans and mice in many chromosomal locations. Despite the difference in the numbers of genes, the numbers of large genomic clusters are nearly the same for humans and mice. These observations suggest that the greater OR gene repertoire in mice has been generated mainly by tandem gene duplication within each genomic cluster.

Keywords: Tandem gene duplication; Birth-and-death process; Genomic cluster; Molecular phylogeny; Comparative genomics

1. Introduction

The mammalian olfactory receptor (OR) gene family contains ~1000 member genes comprising ~2% of the entire set of genes (see Firestein, 2001; Mombaerts, 2004 for review). OR genes are mainly expressed in sensory neurons of olfactory epithelia in nasal cavities and detect thousands of different odor molecules in the environment. They are G-protein-coupled receptors containing seven α-helical transmembrane regions. OR genes have been identified in various vertebrates from lampreys to humans. They are approximately 310 codons long on average, share conserved motifs, and do not have any introns in their coding regions. Because of the large number of member genes, the OR gene family is one of the most informative systems to study the evolutionary dynamics of genomic sequences.

There are several studies of OR genes using the whole genome sequences of humans (Glusman et al., 2001; Zozulya et al., 2001; Niimura and Nei, 2003) or mice (Zhang and Firestein, 2002; Young et al., 2002; Godfrey et al., 2004; Zhang et al., 2004). Recently we reported that there are 802 OR genes in the complete human genome sequences and 52% of them are pseudogenes (Niimura and Nei, 2003). By contrast, mice have ~1400 OR genes, but the fraction of pseudogenes is 20–25% (Zhang and

Abbreviations: OR, olfactory receptor; NJ, neighbor-joining; PC, Poisson correction; kb, kilobase(s); Mb, megabase(s); MHC, major histocompatibility complex.

* Corresponding author. Present address: Department of Bioinformatics, Medical Research Institute, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan. Tel.: +81 3 5803 5840; fax: +81 3 5803 0247.
E-mail address: niimura@bioinfo.tmd.ac.jp (Y. Niimura).

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.09.025

Firestein, 2002; Young et al., 2002; Zhang et al., 2004; present study). Human OR genes are distributed on almost all chromosomes and typically form genomic clusters (Glusman et al., 2001). Evolutionarily closely related genes are usually tandemly arranged in a genomic cluster, but one cluster may contain distantly related genes as well (Niimura and Nei, 2003). These findings suggest that OR gene clusters have been subject to genomic rearrangements. Therefore, it is interesting to compare the chromosomal locations of human OR genes with those of another species to study their evolutionary change. Moreover, comparative analysis of OR genomic clusters between humans and mice will reveal evolutionary processes that have shaped the difference in OR genes between the two species.

The purpose of this paper is to study these problems. To this end, we first identified the genomic locations of all mouse OR genes from the mouse genome sequence by using the same criteria as our previous work for human OR genes (Niimura and Nei, 2003). We then conducted a detailed phylogenetic analysis of human and mouse genes and studied the evolutionary changes of OR functional genes and pseudogenes.

2. Materials and methods

2.1. Detection of OR genes from the mouse genome

Functional and nonfunctional OR genes were detected from the whole mouse genome sequences by conducting homology search. DNA sequences of the mouse genome were downloaded from http://genome.ucsc.edu (mm3, the Feb. 2003 version). We retrieved translated amino acid sequences of 904 mouse OR genes from the DNA Data Bank of Japan, http://www.ddbj.nig.ac.jp (accession numbers, AY072961-AY074256; Zhang and Firestein, 2002). We then performed TBLASN search (Altschul et al., 1997) with the cutoff E-value of 10^{-20} against the whole mouse genome sequences using these 904 mouse OR genes as queries. All of the matches detected by the homology search were regarded as functional or nonfunctional OR genes. The criteria to identify functional genes among these matches were the same as those in our previous work for human ORs (Niimura and Nei, 2003).

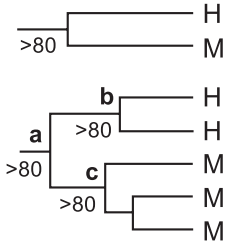


Fig. 1. Identification of orthologous genes between humans and mice. When a phylogenetic clade that contained a mouse gene and a human gene was supported by a >80% bootstrap value, these genes were regarded as orthologous to each other (top). A gene from one species may be orthologous to two or more genes from the other species. As shown in the bottom figure, human genes (H) and mouse genes (M) were regarded as orthologous when a clade (**b**) for the human genes was supported by a >80% bootstrap value, a clade (**c**) for the mouse genes was supported by a >80% bootstrap value, and the larger clade (**a**) including clades **b** and **c** was supported by a >80% bootstrap value. We obtained essentially the same results when we used 70% or 90% bootstrap values instead of 80%.

Table 1
Comparison of OR gene families between mice and humans

	Mouse	Human[a]
Total no. of functional genes and pseudogenes	1391	802 (724)[b]
No. of functional genes	1037	388
No. of pseudogenes	354	414 (336)[b]
Fraction of pseudogenes	25%	52% (46%)[b]
No. of genomic clusters	69	95 (66)[b]
No. of genomic clusters containing ≥5 ORs	34	34 (32)[b]

[a] From Niimura and Nei (2003).
[b] H* pseudogenes are excluded (see Section 4).

2.2. Phylogenetic analysis

The phylogenetic tree in Fig. 3 was constructed in the following way. The amino acid sequences of 1425 (=1037+388) mouse and human functional genes were aligned by the computer program FFT-NS-i (Katoh et al., 2002). Poisson correction (PC) distances (Nei and Kumar, 2000) were calculated using 213 amino acid sites after all alignment gaps were eliminated. A phylogenetic tree was constructed from these distances using the neighbor-joining (NJ) method (Saitou and Nei, 1987) as implemented in the program LINTREE (http://www.bio.psu.edu/People/Faculty/Nei/Lab; Takezaki et al., 1995).

2.3. Orthologous OR genes between mice and humans

We identified orthologous OR genes between mice and humans on the basis of phylogenetic analysis. Almost all of the 40 phylogenetic clades identified in this study included both mouse and human genes (Fig. 3). Therefore, a human gene orthologous to a mouse gene in clade X also belongs to clade X. For this reason, we identified orthologous gene pairs belonging to clade X using a phylogenetic tree that contains only clade X genes and the outgroup. As the outgroup, a class II gene (HsOR1.1.4) was used when clade X is a class I gene clade, and a class I gene (HsOR11.3.2) was used when clade X is one of the class II gene clades. Orthologous genes for unclassified class II genes were identified from the phylogenetic tree in Fig. 3. The method to identify orthologous genes using these phylogenetic trees is illustrated in Fig. 1.

3. Results

3.1. OR genes in the mouse genome

Conducting extensive homology search, we detected 1037 putatively functional OR genes that have intact coding regions and 354 apparent pseudogenes in the mouse genome (Table 1). The total number of functional genes and pseudogenes (1391) are nearly the same as those of previous studies (Zhang and Firestein, 2002; Young et al., 2002; Zhang et al., 2004). The fraction of pseudogenes was approximately 25%. There were four OR gene sequences containing undetermined nucleotides and were regarded as pseudogenes. Of these 1391 OR genes, 12 functional genes and 12 pseudogenes were detected from the unassembled sequences. The precise chromosomal locations were identified for other 1367 functional genes and pseudogenes. The nucleotide and amino acid sequences of the OR genes and their chromosomal locations are available from our web site, http://mep.bio.psu.edu/databases/.

The distribution of OR genes on the mouse chromosomes is shown in Fig. 2. We defined OR genomic clusters using the criterion that any distances between two neighboring OR genes (including pseudogenes) in a cluster are less than 500

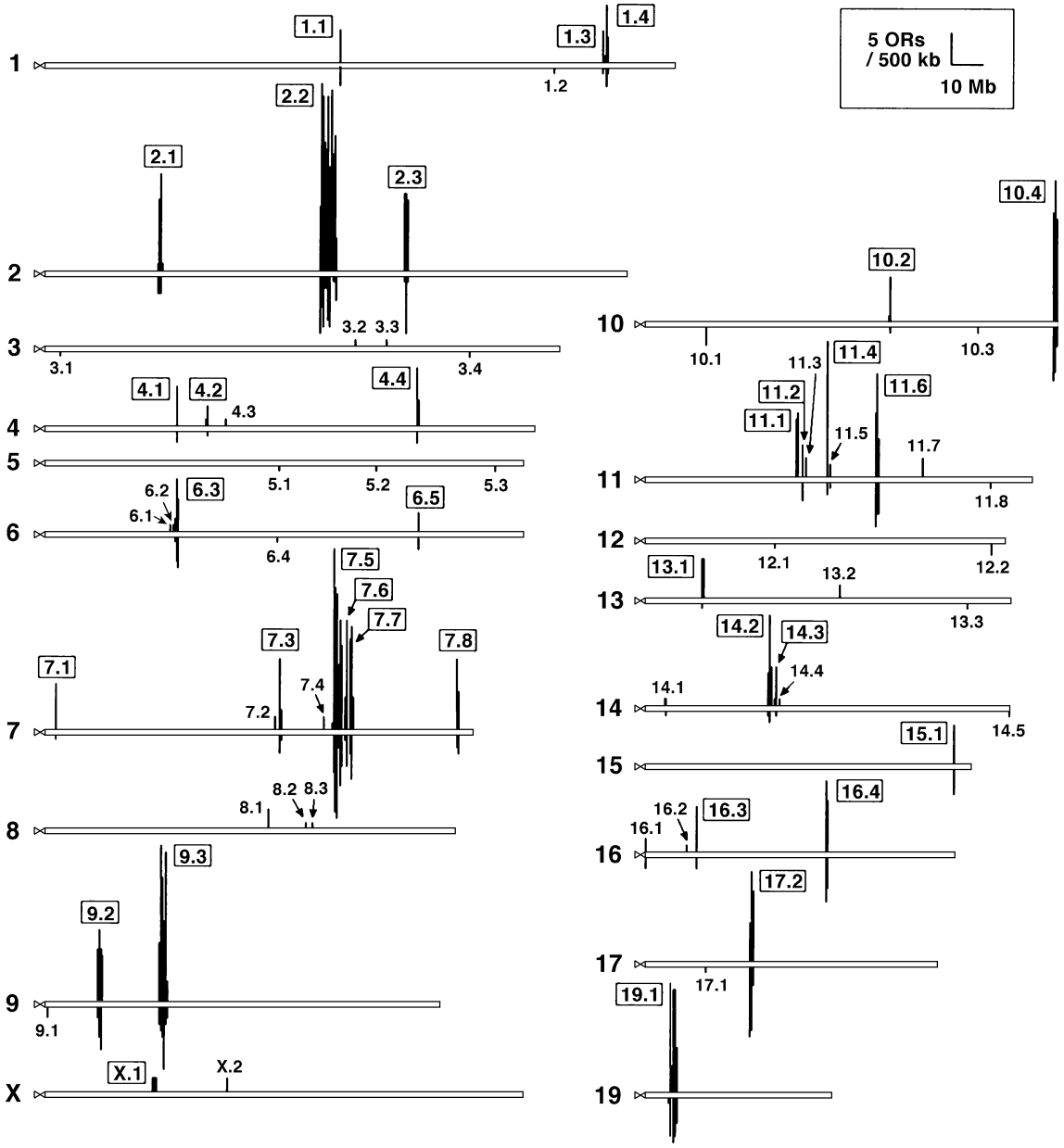
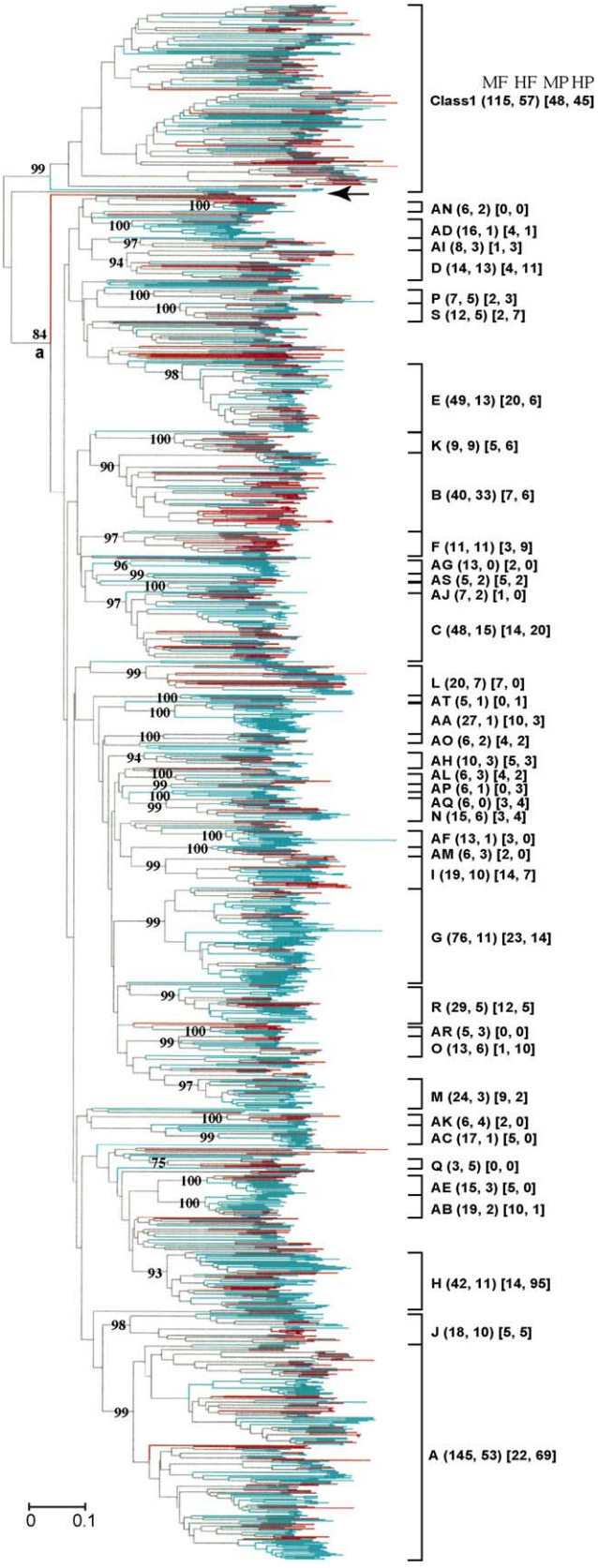


Fig. 2. Distribution of OR genes in the mouse genome. Vertical bars above and below the chromosomes show the locations of functional OR genes and pseudogenes, respectively. The height of each bar indicates the number of OR genes in a non-overlapping 500-kb window at the position. A genomic cluster containing five or more OR genes (including pseudogenes) is shown by a box. Chromosome 18 is omitted, because OR genes were not found on the chromosome. Chromosome Y has not been sequenced.



kb. This criterion is the same as that used in our study of human genes (Niimura and Nei, 2003). Excluding the genes detected from the unassembled sequences, 69 genomic clusters (including singletons) were identified. This number is smaller than that for humans (Table 1). However, the number of genomic clusters containing five or more OR genes is the same between humans and mice (see Section 4). The genomic clusters were named by using chromosome number and a number indicating the order of the cluster in each chromosome (Niimura and Nei, 2003). For example, the first genomic cluster from the centromere on mouse chromosome 1 is Mm1.1. In this paper, Mm and Hs refer to the mouse and human clusters, respectively. The largest cluster, Mm2.2, contains 267 functional OR genes and pseudogenes (19% of all OR genes in mice) and occupies an approximately 5 Mb region. This cluster is comparable to the human major histocompatibility complex (MHC) genomic region (3.6 Mb) containing 224 functional genes and pseudogenes, one of the most gene-rich regions in mammalian genomes (The MHC Sequencing Consortium, 1999). The distribution of the distances between two consecutive OR genes on the same chromosome is nearly the same as that of the human (data not shown; see Fig. 6 in Niimura and Nei, 2003).

3.2. Classification of mouse OR genes

To classify mouse OR genes into phylogenetic clades, we conducted a phylogenetic analysis using 1037 mouse functional genes and 388 human functional genes (Fig. 3). These genes were clearly separated into class I and class II genes (Glusman et al., 2000). Previously, we identified 19 phylogenetic clades (clades A–S) for human class II OR genes (Niimura and Nei, 2003). As shown in Fig. 3, almost all of these clades were supported by >90% bootstrap values, although the number of genes used was more than three times greater than that in the previous analysis. The only clade which had a low bootstrap value (75%) was clade Q. Moreover, we found 20 more class II clades that were

Fig. 3. NJ tree constructed for 1037 mouse functional OR genes and 388 human functional OR genes. Forty phylogenetic clades including class I gene clade were identified in this study. The first and second numbers in parentheses refer to the numbers of functional OR genes from mice and humans, respectively. The first and second numbers in brackets refer to the numbers of OR pseudogenes from mice and humans, respectively. The bootstrap value obtained from 1000 replications is shown at the branch determining each clade. Blue and red lines indicate branches for mouse and human genes, respectively. The scale bar shows the estimated number of amino acid substitutions per site. All the clades previously identified for human genes (class I gene clade and class II gene clades A–S; Niimura and Nei, 2003) were supported by >90% bootstrap values except for clades Q, which was supported by a bootstrap value of only 75%. The definition of clade M was slightly modified from the previous one considering an interior branch supported by a >90% bootstrap value. The arrow sign indicates two mouse genes and one human gene that diverged very early in the evolution of class II OR genes. These three genes form a clade with a 100% bootstrap value (the value is not shown).

supported by >90% bootstrap values and contained five or more mouse OR genes. When a clade or clades were included in another larger clade, we used the larger clade for the classification of OR genes, ignoring the smaller ones. We named these phylogenetic clades AA, AB, . . ., AT from the clade containing the largest number of mouse genes to the smallest. We have therefore identified 39 class II gene clades, but 126 genes have remained unclassified.

Almost all the clades include both mouse and human genes, indicating that the divergence of these clades occurred before the divergence of the two species. The largest clade in mice, clade A, was also the largest clade in human class II genes. (The numbers of mouse and human genes included in each clade are presented in parentheses in Fig. 3). However, the relative size of a phylogenetic clade was not the same for mice and humans. For example, the second largest clade in mice, clade G, was the seventh largest clade in humans. Clade AA contained 27 mouse genes but only one human gene. By contrast, clades B, D, and F contained nearly the same number of genes from mice and humans. Fig. 3 also indicates that two mouse genes and one human gene (shown by the arrow) diverged very early in the evolution of class II genes, because the interior branch **a** in Fig. 3 is supported by a bootstrap value of 84%.

Mouse and human OR pseudogenes were also classified into the clades by phylogenetic analysis as is illustrated in Fig. 4. Phylogenetic trees containing many pseudogenes were not very reliable, because some pseudogenes included deletions of many nucleotides. For this reason, we constructed a phylogenetic tree one by one for each of the 354 mouse and 414 human pseudogenes together with all the mouse and human functional genes. When a pseudogene was included in clade A of the phylogenetic tree, for example, the pseudogene was assigned to clade A (Fig. 4B). The numbers of mouse and human pseudogenes in each clade are shown in brackets in Fig. 3. There was one mouse-specific clade (AG), in which no functional genes and no pseudogenes from humans existed. By contrast, there was no human-specific clade consisting of human genes only. As shown in Fig. 5, the ratio of the number of pseudogenes to that of functional genes was highly variable for human clades (8.6 for clade H to 0 for clade L), while the ratio was much less variable for mouse clades. Seventy-one mouse and 65 human pseudogenes remained unclassified.

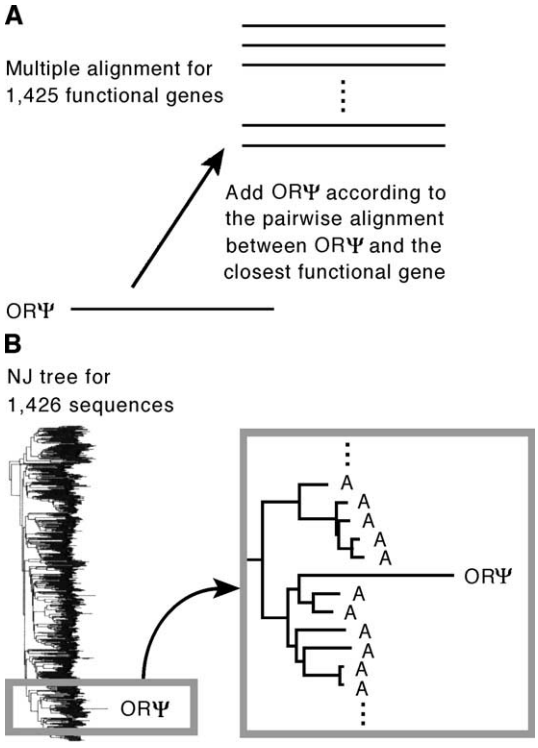


Fig. 4. Classification of pseudogenes into phylogenetic clades. ORψ represents one of the 768 (=354+414) mouse and human pseudogenes. The analysis shown in this figure was conducted for each of the 768 pseudogenes. (A) Multiple alignment containing 1426 sequences (1425 mouse and human functional genes and ORψ) was constructed. For this purpose, we first identified a functional gene that is closest to ORψ by taking the best hit of BLASTP search (Altschul et al., 1997) and constructed an amino acid sequence alignment between ORψ and its closest functional gene using the program CLUSTALW (Thompson et al., 1994). (Here we used CLUSTALW instead of FFT-NS-i, simply because FFT-NS-i is not applicable to a pairwise alignment.) According to the pairwise alignment, the sequence of ORψ was added to the multiple alignment of the 1425 functional genes that was used for constructing the tree in Fig. 3. (B) A NJ phylogenetic tree was constructed for the 1426 sequences. PC distances were calculated between the amino acid sequences of ORψ and each of the 1425 functional genes using the pairwise-deletion option (Nei and Kumar, 2000). For the distances among the functional genes, the same PC distances as those used for constructing the tree in Fig. 3 were used. (We also constructed a tree using the complete-deletion option, but the result of the classification of pseudogenes was essentially the same.) Using the new phylogenetic tree, ORψ was assigned to one of the 40 phylogenetic clades in the following way. If ORψ was included in clade X of the phylogenetic tree, ORψ was assigned to clade X. Otherwise, ORψ remained unclassified. This figure indicates that ORψ is a clade A pseudogene.

3.3. Orthologous relationships of OR genes between mouse and human genomic clusters

The physical maps of OR genes in three mouse genomic clusters are shown in Supplementary data, Fig. S1. This figure indicates that OR genomic clusters in mice have the following features. (i) Functional and nonfunctional OR genes belonging to the same clade tend to form a tandem array in a genomic cluster. (ii) One genomic cluster may contain several clades that are distantly related in the phylogenetic tree. For example, genomic cluster Mm2.2 contains OR genes from 11 clades, A, M, N, O, AD, AF, AM, AO, AP, AQ, and AR, but these clades are dispersed in the phylogenetic tree in Fig. 3. Similar observations have been made with human OR genes (Niimura and Nei, 2003).

Using phylogenetic trees, we examined orthologous gene pairs between mice and humans (see Section 2.3 and Fig. 1). We identified 205 such pairs, which contained 306 mouse genes and 241 human genes. (Note that an orthologous gene pair may contain two or more genes from each species). Fig. 6 shows the orthologous relationships of OR genes

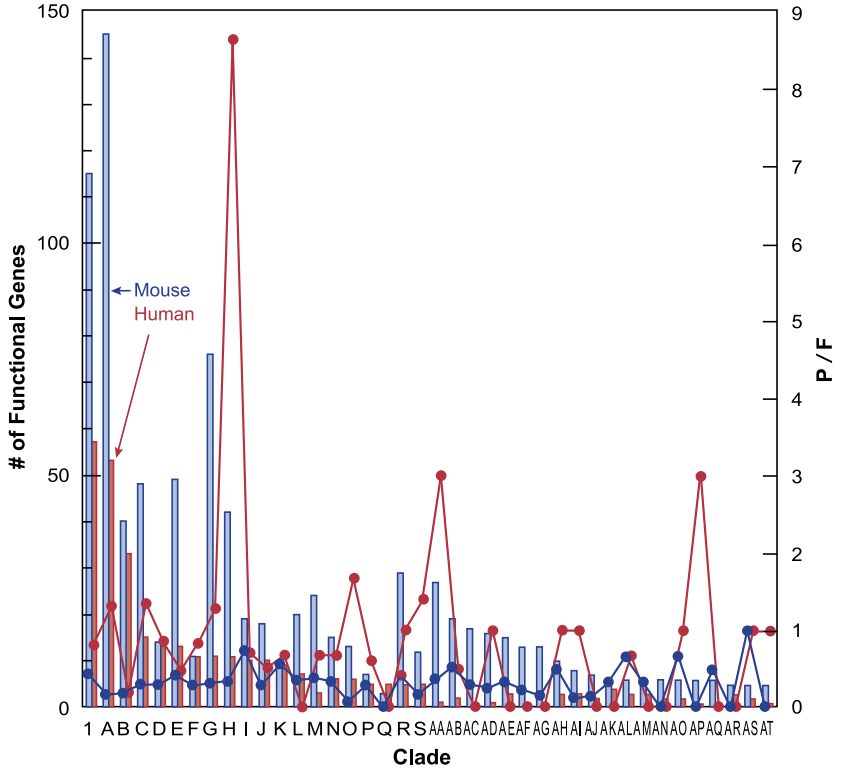


Fig. 5. Number of functional OR genes (histogram, left label) and the ratio of pseudogenes to functional genes (line graph, right label) for different clades. Blue and red represent mouse and human genes, respectively. "1" represents class I genes.

between mouse and human genomic clusters. In general, the gene order and the transcriptional directions of the orthologous genes are well conserved between the two species (Young et al., 2002). Out of 34 mouse genomic clusters with five or more OR genes, 27 clusters were found to contain genes orthologous to human genes (Table 2). Twenty-eight out of thirty-four human clusters with five or more OR genes contained genes orthologous to mouse genes. Orthologous relationships were ambiguous for the genes in most other clusters. However, we found one mouse cluster (Mm4.4) that clearly does not contain any genes orthologous to human genes. This cluster included 15 OR genes (including pseudogenes) and all of them belonged to the mouse-specific clade, AG.

The largest genomic cluster of mice, Mm2.2, contained OR genes orthologous to human genes in four consecutive clusters, Hs11.8–Hs11.11, which were located on the same chromosome with intervals of 1.3–3.3Mb. The cluster Hs11.11 contained all of the 11 clade genes found in the cluster Mm2.2, and the gene order was largely conserved between the two clusters. However, the number of OR genes in Mm2.2 was almost twice as large as the total number of OR genes in the four human clusters, Hs11.8–Hs11.11 (Table 2). There were many other clusters in which the number of orthologous genes was much larger in the mouse than in the human. For example, Mm9.3 contained 118 genes, whereas Hs11.18 contained 43 genes orthologous to Mm9.3 genes (Fig. 6 and Table 2). However, there were also cases in which the number of OR genes was greater in the human cluster than in the mouse cluster (e.g., Mm4.2 vs. Hs9.4). The OR genomic clusters in the two species were not always well conserved. For example, Hs1.5 contained genes orthologous to the mouse genes in five clusters, Mm7.3, Mm11.4, Mm11.5, Mm14.1, and Mm16.3, which were located on four different chromosomes. This suggests that complicated genomic rearrangements have occurred in the past. The orthologous relationships observed between Mm14.2 and Hs14.1 indicates that a genomic block inversion has occurred in either the mouse or the human lineage.

4. Discussion

In this paper, we showed that the organization of OR genomic clusters is generally well conserved between humans and mice. This suggests that many genomic rearrangements occurred in the regions of OR genomic clusters (Niimura and Nei, 2003) before the human-mouse divergence. Moreover, although the number of OR genes is much larger in mice than in humans, the number of genomic clusters that contain five or more OR genes are the same (Table 1, see below). We did not find any clear cases in which the entire genomic cluster is duplicated in mice. These observations indicate that the difference in the

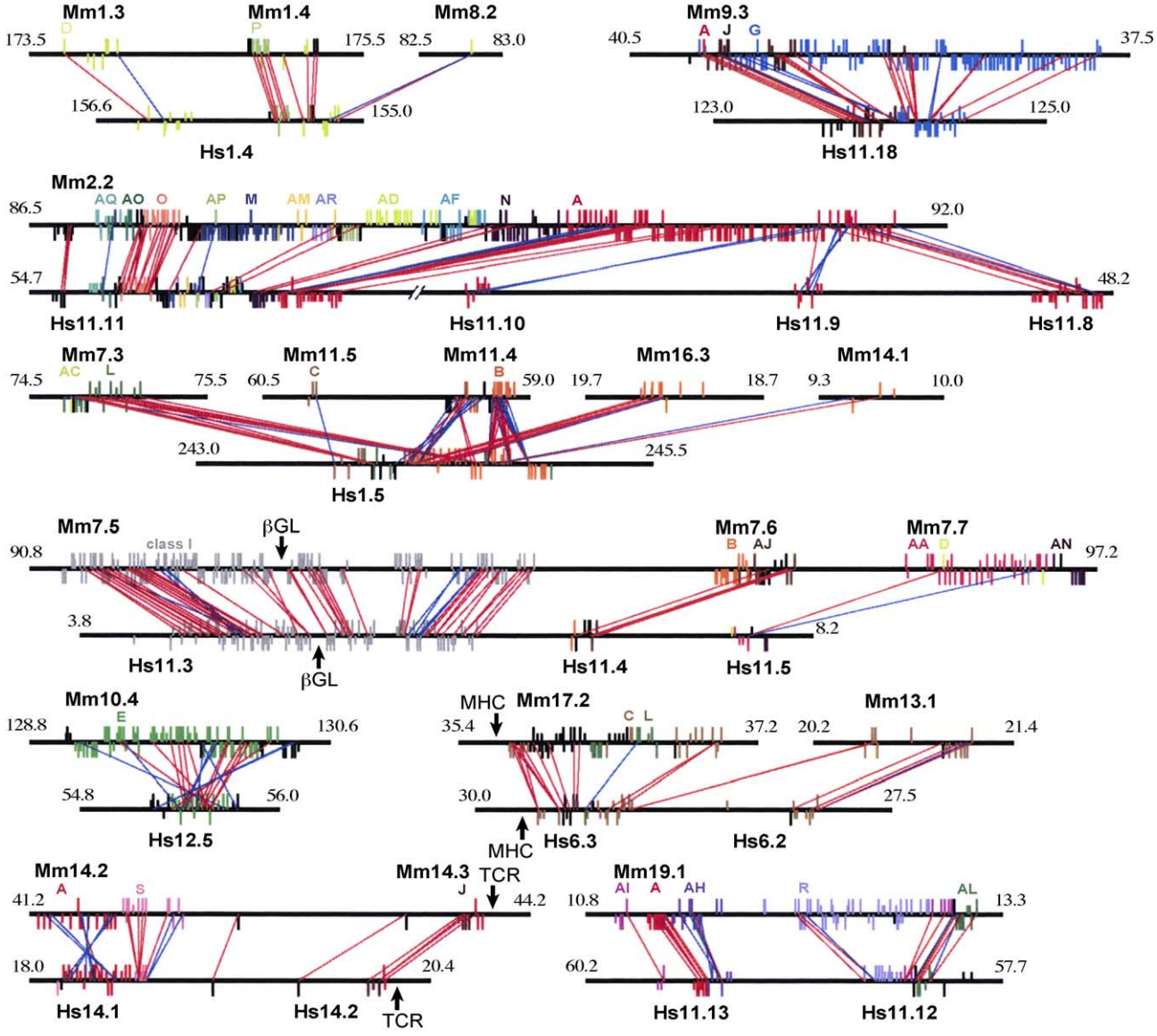


Fig. 6. Orthologous relationships of OR genes between mouse and human genomic clusters. Long and short vertical bars show the locations of functional and nonfunctional OR genes, respectively. A vertical bar above a horizontal line indicates the opposite transcriptional direction to that below a horizontal line. Different colors represent different clades. Unclassified class II OR genes are shown in black. Red and blue lines connecting mouse and human OR genes represent orthologous gene pairs. A red line indicates that transcriptional directions of orthologous genes are conserved between mice and humans, while a blue line indicates that they are inverted. One gene may be orthologous to two or more genes in another species (see Fig. 1). For example, six clade B genes in Mm11.4 are orthologous to three genes in Hs1.5. Numbers at the end of a horizontal line show coordinates in a chromosome in Mb. A 2.5-Mb region between Hs11.10 and Hs11.11 was omitted. Arrows show the locations of the β-globin gene clusters (βGL), MHC class I regions (MHC), and T-cell receptor α/δ loci (TCR).

number of OR genes between humans and mice have been generated mainly by repeated tandem gene duplication within each genomic cluster.

Previously we reported that all the human class I OR genes (including pseudogenes) are located in one cluster, Hs11.3, and this cluster does not contain any class II genes (Niimura and Nei, 2003). The situation is almost the same for mice. The mouse cluster Mm7.5 did not contain any class II genes, and almost all class I genes were found in this cluster (Fig. 6). The exceptions are two functional class I genes found in Mm7.7 and one class I pseudogene in Mm11.8. Moreover, one functional and one nonfunctional class I genes were found from the unassembled sequences. However, the possibility of assembly error cannot be excluded at present. The reason why the class I genes are primarily conserved in given regions of chromosomes is unknown.

Some of the orthologous relationships of mouse and human OR genomic clusters (Table 2) have previously been reported. Both of the class I clusters, Mm7.5 and Hs11.3, include the β-globin gene cluster (Fig. 6; Bulger et al., 1999, 2000). The genomic clusters Mm17.2 and Hs6.3 are located

Table 2
Mouse and human OR genomic clusters containing orthologous genes

Mouse cluster name	No. of OR genes	Human cluster name[a]	No. of OR genes[a]	Clade[b]	M/H[c]
Mm1.1	8	Hs2.4	3	Un	2.7
Mm1.3, Mm1.4, Mm8.2	7+17+1	Hs1.4	28	D, P, Un	0.9
Mm2.1	37	Hs9.6	15	Q, AE, Un	2.5
Mm2.2	267	Hs11.8, Hs11.9, Hs11.10, Hs11.11	21+8+9+98	A, M, N, O, AD, AF, AM, AO, AP, AQ, AR, Un	2.0
Mm2.3	46	Hs1.1, Hs5.4, Hs8.1, Hs15.2	5+5+1+8	A, B[d], C[d]	2.4
Mm4.1	8	Hs9.1	7	F	1.1
Mm4.2	5	Hs9.4	12	F	0.4
Mm6.3	25	Hs7.6	23	K, AS, Un	1.1
Mm6.5	5	Hs10.2	2	AC, Un	2.5
Mm7.3, Mm11.4, Mm11.5, Mm14.1, Mm16.3	18+23+3+3+9	Hs1.5	50	B, C, L, AC[e], Un	1.1
Mm7.5	158	Hs11.3	102	class1	1.5
Mm7.6	28	Hs11.4	8	B, AJ, Un	3.5
Mm7.7	45	Hs11.5	8	class1[e], D, AA, AN, Un	5.6
Mm8.1, Mm9.2	3+48	Hs19.2	17	B, H, Un	3.0
Mm9.3	118	Hs11.18	43	A, G, J, Un	2.7
Mm10.2	9	Hs19.3	13	D[d], H, Un	0.7
Mm10.4	74	Hs12.5	26	E, Un	2.8
Mm11.6	49	Hs17.1	16	AB, AK, Un	3.1
Mm13.1, Mm17.2	13+55	Hs6.2, Hs6.3	9+26	C, L, Un	1.9
Mm14.2	28	Hs14.1	30	A, S, Un	0.9
Mm14.3	8	Hs14.2	6	A, J, Un	1.3
Mm16.4	31	Hs3.3	17	I	1.8
Mm19.1	84	Hs11.12, Hs11.13	23+15	A, R, AH, AI, AL, Un	2.2

[a] From Niimura and Nei (2003).
[b] "Un" represents unclassified class II genes.
[c] Ratio of the number of OR genes in the mouse cluster to that in the human cluster.
[d] Found only in the human cluster.
[e] Found only in the mouse cluster.

at the end of the MHC class I regions in both species (Younger et al., 2001; Amadou et al., 2003). The clusters Mm14.3 and Hs14.2 are located next to the Vα gene segment region of the T-cell receptor α/δ loci (Lane et al., 2002). Furthermore, the orthologous relationships between the following genomic clusters have been reported: Mm4.1 vs. Hs9.1 and Mm4.2 vs. Hs9.2 (Hoppe et al., 2003), Mm11.6 vs. Hs17.1 (Lapidot et al., 2001), and Mm7.6 vs. Hs11.4 (Lane et al., 2001). However, the conclusions derived from a comparison of a particular pair of genomic clusters between mice and humans cannot be generalized to the entire OR gene family, because the ratio of the number of OR genes in a mouse genomic cluster to that in the human cluster having orthologous relationships is quite variable among different clusters (Table 2).

In Fig. 5, clade H showed by far the highest ratio of pseudogenes to functional genes. This observation is explained in the following way. As we mentioned earlier, we constructed a phylogenetic tree for a given pseudogene and all the functional genes in humans and mice to assign the pseudogene to one of the phylogenetic clades. We found that 78 out of the 95 clade H pseudogenes in humans formed clades with a single human functional gene, HsOR19.4.14, that belongs to clade H. Here we call these 78 pseudogenes H* pseudogenes. Therefore, most of the clade H pseudogenes in humans were actually H* pseudogenes. It has been reported that OR genes belonging to the OR7E subfamily as identified by Glusman et al. (2000) have expanded in the human genome as a part of large segmental duplications, duplicated blocks of genomic DNA that contain many repetitive elements (Newman and Trask, 2003). Our H* pseudogenes belong to this OR7E subfamily. Phylogenetic analysis of the H* pseudogenes suggested that the increase of H* genes occurred mostly after they became pseudogenes (data not shown), indicating that they have not contributed to the functional diversification of human OR genes. By contrast, there are no such pseudogenes in mice. (At most four mouse pseudogenes formed clades with the same mouse functional gene). We can therefore exclude these H* pseudogenes in the comparison of OR gene families between humans and mice. If we do this, the fraction of pseudogenes in humans reduces to 46% (Table 1). Interestingly, 29 out of the 95 human genomic clusters contain only H* pseudogenes, and two of the 29 clusters contain five H* pseudogenes. If the clusters containing only H* pseudogenes are excluded, the numbers of OR genomic clusters in the human and mouse genomes are nearly the same (Table 1). Therefore, the more dispersed distribution of OR genes in humans than in mice (Young et al., 2002) can be explained by the presence of H* pseudogenes.

The phylogenetic tree in Fig. 3 indicates that the OR multigene family is mainly subject to the birth-and-death model of evolution rather than to concerted evolution, because many phylogenetic clades contain both human and mouse genes. In the birth-and-death model, new genes are created by repeated gene duplication, and some of them acquire a new function and remain in the genome for a long time, while others become pseudogenes or are deleted (Nei, 1969; Nei et al., 1997). By contrast, the model of concerted evolution proposes that member genes in a multigene family are homogenized by gene conversion or unequal crossing-over, predicting higher sequence similarity of genes within species than between species (Smith, 1974). Moreover, as we mentioned earlier, a genomic cluster often contains phylogenetically distantly related genes, which is also incompatible to concerted evolution. Sharon et al. (1999) proposed that OR genes have undergone gene conversion events by analyzing a particular OR genomic cluster in humans and other primates. However, for the reasons mentioned above, the contribution of gene conversion to the evolution of OR genes seems to be quite small.

Acknowledgements

This work was supported by National Institutes of Health Grant GM20293 (to M.N.).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at 10.1016/j.gene.2004.09.025.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Amadou, C., Younger, R.M., Sims, S., Matthews, L.H., Rogers, J., Kumanovics, A., Ziegler, A., Beck, S., Lindahl, K.F., 2003. Co-duplication of olfactory receptor and MHC class I genes in the mouse major histocompatibility complex. Hum. Mol. Genet. 12, 3025–3040.

Bulger, M., van Doorninck, J.H., Saitoh, N., Telling, A., Farrell, C., Bender, M.A., Felsenfeld, G., Axel, R., Groudine, M., von Doorninck, J.H., 1999. Conservation of sequence and structure flanking the mouse and human β-globin loci: the β-globin genes are embedded within an array of odorant receptor genes. Proc. Natl. Acad. Sci. U. S. A. 96, 5129–5134.

Bulger, M., Bender, M.A., van Doorninck, J.H., Wertman, B., Farrell, C.M., Felsenfeld, G., Groudine, M., Hardison, R., 2000. Comparative structural and functional analysis of the olfactory receptor genes flanking the human and mouse β-globin gene clusters. Proc. Natl. Acad. Sci. U. S. A. 97, 14560–14565.

Firestein, S., 2001. How the olfactory system makes sense of scents. Nature 413, 211–218.

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

Glusman, G., Yanai, I., Rubin, I., Lancet, D., 2001. The complete human olfactory subgenome. Genome Res. 11, 685–702.

Godfrey, P.A., Malnic, B., Buck, L.B., 2004. The mouse olfactory receptor gene family. Proc. Natl. Acad. Sci. U. S. A. 101, 2156–2161.

Hoppe, R., Breer, H., Strotmann, J., 2003. Organization and evolutionary relatedness of OR37 olfactory receptor genes in mouse and human. Genomics 82, 355–364.

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066.

Lane, R.P., Cutforth, T., Young, J., Athanasiou, M., Friedman, C., Rowen, L., Evans, G., Axel, R., Hood, L., Trask, B.J., 2001. Genomic analysis of orthologous mouse and human olfactory receptor loci. Proc. Natl. Acad. Sci. U. S. A. 98, 7390–7395.

Lane, R.P., Roach, J.C., Lee, I.Y., Boysen, C., Smit, A., Trask, B.J., Hood, L., 2002. Genomic analysis of the olfactory receptor region of the mouse and human T-cell receptor α/δ loci. Genome Res. 12, 81–87.

Lapidot, M., Pilpel, Y., Gilad, Y., Falcovitz, A., Sharon, D., Haaf, T., Lancet, D., 2001. Mouse-human orthology relationships in an olfactory receptor gene cluster. Genomics 71, 296–306.

The MHC Sequencing Consortium, 1999. Complete sequence and gene map of a human major histocompatibility complex. Nature 401, 921–923.

Mombaerts, P., 2004. Genes and ligands for odorant, vomeronasal and taste receptors. Nat. Rev., Neurosci. 5, 263–278.

Nei, M., 1969. Gene duplication and nucleotide substitution in evolution. Nature 221, 40–42.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Nei, M., Gu, X., Sitnikova, T., 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc. Natl. Acad. Sci. U. S. A. 94, 7799–7806.

Newman, T., Trask, B.J., 2003. Complex evolution of 7E olfactory receptor genes in segmental duplications. Genome Res. 13, 781–793.

Niimura, Y., Nei, M., 2003. Evolution of olfactory receptor genes in the human genome. Proc. Natl. Acad. Sci. U. S. A. 100, 12235–12240.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

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

Smith, G.P., 1974. Unequal crossover and the evolution of multigene families. Cold Spring Harbor Symp. Quant. Biol. 38, 507–513.

Takezaki, N., Rzhetsky, A., Nei, M., 1995. Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. 12, 823–833.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Young, J.M., Friedman, C., Williams, E.M., Ross, J.A., Tonnes-Priddy, L., Trask, B.J., 2002. Different evolutionary processes shaped the mouse and human olfactory receptor gene families. Hum. Mol. Genet. 11, 535–546.

Younger, R.M., Amadou, C., Bethel, G., Ehlers, A., Lindahl, K.F., Forbes, S., Horton, R., Milne, S., Mungall, A.J., Trowsdale, J., Volz, A., Ziegler, A., Beck, S., 2001. Characterization of clustered MHC-linked olfactory receptor genes in human and mouse. Genome Res. 11, 519–530.

Zhang, X., Firestein, S., 2002. The olfactory receptor gene superfamily of the mouse. Nat. Neurosci. 5, 124–133.

Zhang, X., Rodriguez, I., Mombaerts, P., Firestein, S., 2004. Odorant and vomeronasal receptor genes in two mouse genome assemblies. Genomics 83, 802–811.

Zozulya, S., Echeverri, F., Nguyen, T., 2001. The human olfactory receptor repertoire. Genome Biol. 2, research0018.1–0018.12.