

Similar Numbers but Different Repertoires of Olfactory Receptor Genes in Humans and Chimpanzees

Yasuhiro Go*†‡ and Yoshihito Niimura§

*Department of Organismic and Evolutionary Biology, Harvard University; †Department of Biosystems Science, Graduate University for Advanced Studies (Sokendai), Hayama, Kanagawa, Japan; ‡Laboratory for Biodiversity (Global COE Project), Graduate School of Science, Kyoto University, Inuyama, Aichi, Japan; and §Department of Bioinformatics, Medical Research Institute, Tokyo Medical and Dental University, Bunkyo-ku, Tokyo, Japan

Animals recognize their external world through the detection of tens of thousands of chemical odorants. Olfactory receptor (OR) genes encode proteins for detecting odorant molecules and form the largest multigene family in mammals. It is known that humans have fewer OR genes and a higher fraction of OR pseudogenes than mice or dogs. To investigate whether these features are human specific or common to all higher primates, we identified nearly complete sets of OR genes from the chimpanzee and macaque genomes and compared them with the human OR genes. In contrast to previous studies, here we show that the number of OR genes (~810) and the fraction of pseudogenes (51%) in chimpanzees are very similar to those in humans, though macaques have considerably fewer OR genes. The pseudogenization rates and the numbers of genes affected by positive selection are also similar between humans and chimpanzees. Moreover, the most recent common ancestor between humans and chimpanzees had a larger number of functional OR genes (>500) and a lower fraction of pseudogenes (41%) than its descendants, suggesting that the OR gene repertoires are in a phase of deterioration in both lineages. Interestingly, despite the close evolutionary relationship between the 2 species, approximately 25% of their functional gene repertoires are species specific due to massive gene losses. These findings suggest that the tempo of evolution of OR genes is similar between humans and chimpanzees, but the OR gene repertoires are quite different between them. This difference might be responsible for the species-specific ability of odor perception.

Introduction

Olfaction, the sense of smell, is essential for the survival of most animals. Versatile odorant molecules in the environment are detected by olfactory receptors (ORs) that are mainly expressed in sensory neurons of the main olfactory epithelium. It is known that OR genes form a very large multigene family, and there are ~1,000 OR genes in mammalian genomes (Buck and Axel 1991; Mombaerts 2004; Niimura and Nei 2006). They are members of G-protein-coupled receptors and have no introns in their coding regions.

Due to recent progress in whole-genome sequencing, the entire repertoires of OR genes from various species have been identified, including humans (Glusman et al. 2001; Zozulya et al. 2001; Niimura and Nei 2003), mice (Young et al. 2002; Zhang and Firestein 2002; Godfrey et al. 2004; Zhang et al. 2004; Niimura and Nei 2005a), dogs (Quignon et al. 2003, 2005; Olender et al. 2004; Niimura and Nei 2007), and other vertebrates (Niimura and Nei 2005b, 2007). These studies have revealed wide variations in the number of OR genes among different species. The number of functional OR genes ranges from <100 in some fishes and <400 in humans to >1,200 in rats and the fraction of OR pseudogenes ranges from <20% in opossums and ~25% in mice or dogs to >50% in humans or platypuses (Niimura and Nei 2005b, 2007).

It is generally thought that higher primates are vision-oriented animals and that their olfactory abilities have therefore retrogressed. Gilad et al. (2004) sequenced ~100 OR genes that were randomly chosen from each of 19 primate species, including apes, Old World Monkeys (OWMs), New World Monkeys (NWMs), and prosimians (see Gilad et al.

[2007] for correction). They found that the fractions of OR pseudogenes in apes and OWMs are significantly higher than those in most NWMs and prosimians. Interestingly, there was one exceptional NWM species, the howler monkey, whose pseudogene fraction was as high as those of apes and OWMs. From these observations, they hypothesized that the loss of OR genes in primates coincides with the acquisition of full trichromatic vision. In the same paper, they also argued that the fraction of OR pseudogenes in humans is even much higher than that of apes or OWMs. In another paper (Gilad et al. 2003), the same group sequenced 50 orthologous OR genes from chimpanzees, gorillas, and orangutans and compared them with human OR genes, reporting that the accumulation of mutations disrupting the coding regions is much faster in humans than in chimpanzees and other apes. In Gilad et al. (2005), they identified 899 OR gene candidates from the first version of the chimpanzee genome sequences (4× coverage). By analyzing these data, they again concluded that humans have a significantly higher fraction of OR pseudogenes than chimpanzees. Therefore, these studies have suggested that the OR gene repertoire has been declining particularly in the human lineage after the separation from our closest relative, chimpanzees.

In this paper, we ask whether or not the extent of the deterioration of OR genes is truly higher in humans than in chimpanzees. This is an important question for understanding human-specific characteristics and the evolution of the sensory system in primates. For this purpose, we used an improved version of the chimpanzee genome sequences with 6× coverage and macaque OR genes as outgroups for human–chimpanzee comparisons. As a result, in contrast to the previous assertion, we did not find any evidence of significant differences between humans and chimpanzees in the number of functional OR genes, the fraction of OR pseudogenes, or the extent of positive selection. We did, however, find that the repertoires of OR genes differ substantially between the 2 species.

Key words: olfactory receptor gene, human evolution, chimpanzee, macaque, multigene family, gene gain and loss.

E-mail: yasuhirigo@gmail.com.

Mol. Biol. Evol. 25(9):1897–1907. 2008

doi:10.1093/molbev/msn135

Advance Access publication June 17, 2008

Materials and Methods

OR Gene Identification

The assembled genome sequences of chimpanzees (panTro2, released in March 2006) and macaques (rheMac2, released in January 2006) were retrieved from the FTP server of the University of California Santa Cruz (UCSC) Genome Browser (<http://genome.ucsc.edu>). The method to identify functional OR genes and pseudogenes was described in detail in a previous paper (Niimura and Nei 2007). We used human OR genes that were identified previously (Niimura and Nei 2003, 2007). The nucleotide sequences of chimpanzee and macaque OR genes are provided in supplementary data sets 1 and 2, respectively (Supplementary Material online).

H* pseudogenes (see Results) were identified by a method similar to that in Niimura and Nei (2005a). We conducted BlastP searches (Altschul et al. 1997) against all functional OR genes in humans by using translated amino acid sequences of OR pseudogenes in humans, chimpanzees, and macaques as queries. We defined a pseudogene that best hits to *HsOR19.2.14* as an H* pseudogene. H* pseudogenes identified by this method are almost identical to 7E pseudogenes in Gilad et al. (2005) and those in the Human Olfactory Receptor Data Exploratorium (HORDE), Build 42 (<http://biportal.weizmann.ac.il/HORDE/>).

The chimpanzee OR genes identified in this study were compared with the previous data. The sequences of 100 and 899 chimpanzee OR genes in Gilad et al. (2004, 2005, 2007) were obtained from the Gilad laboratory Web site (<http://giladlab.uchicago.edu/>) and HORDE, Build 42, respectively. Each of the 100 sequences in Gilad et al. (2004) was compared with the translated amino acid sequences in Gilad et al. (2005) and those in this study by BlastX searches (Altschul et al. 1997), and the gene showing the highest sequence identity was identified. When the nucleotide sequence identity between 2 sequences was $>98\%$, they were regarded to be the same gene.

Identification of Orthologous Gene Sets

Orthologous OR gene sets among humans, chimpanzees, and macaques were determined in the following way.

- (i) To identify orthologous relationships, we first constructed a phylogenetic tree using OR genes that are ≥ 600 bp long. There are 2,038 such sequences in total (749, 730, and 559 sequence for humans, chimpanzees, and macaques, respectively). A multiple sequence alignment was made from these sequences together with an outgroup sequence (a zebrafish OR gene; GenBank accession number AF283560) by ClustalW (Thompson et al. 1994) using a GAOPEN parameter of 20 and GAPEXT of 8. From the alignment, we reconstructed a phylogenetic tree by MEGA3 (Kumar et al. 2004) using the Neighbor-Joining method (Saitou and Nei 1987). The nucleotide *p* distance and the pairwise deletion option were used. From the tree obtained, we identified monophyletic clades containing genes from at least 2 species among humans, chimpanzees, and macaques as candidate orthologous gene sets. The genes contained in each of the clades were aligned by ClustalW. When the nucleotide sequence identity between human and chimpanzee genes in a clade was $\geq 96\%$, the genes were regarded as orthologous. Similarly, when the nucleotide identity between human and macaque genes or that between chimpanzee and macaque genes in a clade was $\geq 90\%$, they were regarded as orthologous. Almost all the nucleotide identities for human–chimpanzee and human–macaque comparisons were $\geq 96\%$ and $\geq 90\%$, respectively, after excluding outliers; for this reason, these cutoff values were chosen. Most ($>90\%$) of the phylogenetic clades for orthologous gene sets identified here were supported with a 100% bootstrap value.
- (ii) The sequences that were <600 bp long (53, 83, and 47 sequences for humans, chimpanzees, and macaques, respectively) were treated in the following way. By using the 53 human sequences as queries, we conducted BlastN searches (Altschul et al. 1997) against all chimpanzee and macaque OR genes. A given query sequence and the best-hit sequence were aligned by ClustalW, and the nucleotide sequence identity between them was calculated. When the nucleotide identity was $\geq 96\%$, they were assumed to be orthologous to each other. Similarly, when the nucleotide identity between the best-hit sequence in macaques and the human query sequence was $\geq 90\%$, they were assumed to be orthologous. Moreover, by using the 83 chimpanzee sequences and the 47 macaque sequences as queries, BlastN searches were carried out against all human and macaque OR genes and against all human and chimpanzee OR genes, respectively. Orthologous genes were identified by using the same cutoff values of nucleotide sequence identity as above (96% for human–chimpanzee orthologs and 90% for human–macaque and chimpanzee–macaque orthologs).
- (iii) After processes (i) and (ii), there were several sequences (41, 71, and 91 sequences for humans, chimpanzees, and macaques, respectively) for which no orthologous genes were identified. These sequences may contain paralogous genes that were generated by lineage-specific gene duplications. To identify such sequences, we conducted intraspecies BlastN searches. By using the 41 human OR genes, self-against-self BlastN searches were performed. A given query sequence and the best-hit sequence (excluding the query sequence itself) were aligned by ClustalW, and the nucleotide sequence identity between them was calculated as above. When the nucleotide identity between the query and the best-hit sequences was $\geq 96\%$, they were assumed to have been generated by gene duplication in the human lineage after the divergence from chimpanzees. When the nucleotide identity was between 90% and 96%, it was assumed that gene duplication occurred after the divergence from macaques and before the divergence from chimpanzees. Furthermore, we calculated the nucleotide sequence identity between the paralogous gene sets identified above

and merged them as genes generated by lineage-specific multiplication using the same criteria of nucleotide identity. We also conducted self-against-self BlastN searches for the chimpanzee sequences and identified paralogous genes in the same manner as human genes. As for macaque sequences, when the nucleotide identity between the query and the best-hit sequences was $\geq 90\%$, they were assumed to have been generated by macaque-specific gene duplication.

All orthologous relationships of OR genes are listed in supplementary table 1 (Supplementary Material online). The ω values for orthologous OR genes shown in supplementary figure 1 (Supplementary Material online) were calculated by the modified Nei–Gojobori method (Zhang et al. 1998) with the transition/transversion ratio of 2.0.

Analysis of Gene Gains and Gene Losses

The number of OR genes in each ancestral node and those of gene gains and gene losses in each lineage were estimated by using the orthologous relationships determined above (supplementary table 1, Supplementary Material online).

To estimate the numbers of functional OR genes and pseudogenes in the most recent common ancestor (MRCA) of humans and chimpanzees, we inferred whether each of the OR genes in the MRCA was a functional gene or a pseudogene in the following way. When at least one of the human and chimpanzee orthologous genes was functional, the MRCA gene was assumed to be functional. Even when both the human and chimpanzee orthologous genes were pseudogenes, if they did not share any disruptive mutations, the gene in the MRCA was assumed to be functional. On the other hand, when both the human and chimpanzee orthologous genes were pseudogenes and they shared at least one disruptive mutation, we assumed that the MRCA gene was a pseudogene. When a pseudogene in one species did not have an orthologous counterpart in the other species, we regarded the MRCA gene as a pseudogene for simplicity. Note that this assumption may cause the fraction of pseudogenes in the MRCA to be overestimated. However, this does not affect the conclusion that the MRCA has a lower fraction of pseudogenes than the extant humans or chimpanzees (see Results).

The probability of the occurrence of i independently pseudogenized orthologous pseudogenes (IPOPs; see Results) under the assumption that pseudogenization events occurred randomly in the human and chimpanzee lineages is given by the hypergeometric distribution

$$P(i) = \frac{\binom{n}{i} \binom{N-n}{m-i}}{\binom{N}{m}} \quad (0 \leq i \leq m),$$

where N (510) is the number of functional genes in the MRCA and n (113) and m (109) are the numbers of MRCA genes that became pseudogenes in the human and chimpanzee lineages, respectively. The expectation of i is calculated as mn/N .

Selection Test

To identify the genes that are subject to lineage-specific positive or negative selection, we conducted maximum likelihood analyses using PAML (Yang and Nielsen 2000; Yang et al. 2000). Orthologous OR gene sets among the 3 species were chosen in the following way. In our data set, there are 266 one-to-one orthologous pairs of functional OR genes between humans and chimpanzees (see table 2). Among them, 219 pairs have orthologous genes in macaques, whether they are functional or not. There were several cases in which 2 or more orthologous genes were present in macaques because of macaque-specific gene duplication. In these cases, when functional genes were present in the macaque orthologs, all functional genes were used; otherwise, all pseudogenes were used.

For each of the orthologous gene sets identified above, we applied 2 kinds of statistical tests. In the branch test, 5 different models were examined. In Model I, the same ω value is assumed along all 3 branches. Model II allows a different ω value in the human lineage from the other lineages. Similarly, Models III and IV allow a different ω value in the chimpanzee and macaque lineages, respectively. Model V allows different ω values in the 3 lineages. The best of the 5 models was determined in the following way. Because the numbers of parameters in Models II, III, and IV are the same, the model showing the highest likelihood score among the 3 models was chosen first and hereafter is called Model X. We then conducted a likelihood ratio test to compare Model X (the alternative hypothesis) with Model I (the null hypothesis) by using χ^2 distribution with 1 degree of freedom (df). When Model I was not rejected, it was assumed to be the best model. Otherwise, a similar statistical test was performed to compare Model V (the alternative hypothesis) with Model X (the null hypothesis) by χ^2 distribution with 1 df. When Model X was rejected, Model V was regarded as the best model. Otherwise, Model X was chosen as the best model.

In the branch-site test, 3 different models were considered (Zhang et al. 2005). In Model I, all codon sites along all branches were assumed to be under negative ($\omega_0 < 1$) or neutral ($\omega_1 = 1$) selection. The fractions of codon sites under negative and neutral selection are p_0 and p_1 , respectively. In Model II, some codons (p_H) in the human lineage were allowed to evolve under positive selection ($\omega_H > 1$). Similarly, in Model III, some codons (p_C) in the chimpanzee lineage were allowed to evolve under positive selection ($\omega_C > 1$). The model that showed a higher likelihood score between Models II and III was chosen and that model (the alternative hypothesis) was compared with Model I (the null hypothesis) by using χ^2 distribution with 2 df.

Results

OR Genes in Humans, Chimpanzees, and Macaques

Table 1 shows the numbers of OR genes and their chromosomal locations in humans, chimpanzees, and macaques. The numbers of functional OR genes and OR pseudogenes in humans and chimpanzees are similar to each other, although macaques have $\sim 25\%$ fewer OR genes than humans and chimpanzees (Niimura and Nei 2007). In this

Table 1
Comparison of OR Gene Repertoires among 3 Primate Species

	Human	Chimpanzee	Macaque
Number of OR genes	802	813	606
Functional genes	387	380	309
Class I	58 (15.0%) ^a	64 (16.8%) ^a	36 (11.7%) ^a
Class II	329	316	273
Truncated genes	0	19	17
Pseudogenes	415	414	280
H* pseudogenes	84	85	8
Pseudogenes and truncated genes (%)	51.7%	53.3%	49.0%
Pseudogenes (%)	51.7%	50.9%	46.2%
Pseudogenes without H* (%)	46.1%	45.2%	45.5%
	Syntenic chromosome		Number of OR genes ^b
	Human	Chimpanzee	Macaque
1	1	1	59 (27)
2p	2a	13	2 (4)
2q	2b	12	0 (4)
3	3	2	10 (20)
4	4	5	0 (10)
5	5	6	4 (4)
6	6	4	16 (22)
7	7	3	14 (19)
8	8	8	1 (8)
9	9	15	25 (14)
10	10	9	1 (4)
11	11	14	172 (180)
12	12	11	18 (21)
13	13	17	0 (8)
14	14	7	22 (20)
15	15	7	6 (12)
16	16	20	2 (1)
17	17	16	13 (5)
18	18	18	0 (2)
19	19	19	20 (21)
20	20	10	—
21	21	3	0 (3)
22	22	10	1 (0)
X	X	X	1 (6)
Y	Y	Y	—
Un	Un	Un	9 (15)

^a Fraction of Class I genes among all functional genes is shown in parentheses.

^b Number of functional genes and pseudogenes including truncated genes (in parentheses) are shown. —, not found.

study, we identified 813 OR genes, including 380 putatively functional genes, in chimpanzees. Mammalian OR genes are known to be classified into Class I and Class II by sequence similarities (Glusman et al. 2000), and the numbers of Class I or Class II genes are similar between humans and chimpanzees. Table 1 also indicates that the fractions of OR pseudogenes are similar among the 3 species (~50%). To compare the fractions of pseudogenes among different species, some caution is necessary because a functional gene embedded in a low-quality genome sequence may appear to be a pseudogene. We therefore separately counted the truncated genes that were at the contig ends and that did not contain any nonsense or frameshift mutations. The truncated genes could become functional upon completion of the genome sequence. However, the estimated fractions of pseudogenes did not substantially change even when all truncated genes were assumed to be functional or were assumed to pseudogenes (table 1).

Gilad et al. (2005) identified 899 OR gene candidates, including 353 intact genes, from the chimpanzee genome sequences. We call this set of 899 genes the Gilad2005 data. Those authors used the previous version of the genome sequences based on the 4× shotgun assembly (NCBI Build 1.1 released in November 2003), whereas in this study we used the 6× shotgun assembly data. Therefore, the quality of the set of OR genes is expected to be improved. In fact, the 353 intact genes in the Gilad2005 data contained a considerable number of truncated genes because they defined an intact gene as a sequence that does not contain any interrupting nonsense or frameshift mutations. On the other hand, in our criteria, a functional gene was defined as a sequence that has the initiation codon and a stop codon at proper positions and does not contain any nonsense or frameshift mutations or long deletions (Niimura and Nei 2007). If we apply our criteria to the 353 intact genes in the Gilad2005 data, the number of functional genes falls to only 269.

To examine the quality of the OR genes identified here in more detail, we compared our data and the Gilad2005 data with the 100 genes in Gilad et al. (2004, 2007) that were experimentally confirmed. We call this set of 100 genes the Gilad2004 data. We found that 19 out of the 100 genes in the Gilad2004 data were not present in the Gilad2005 data, whereas only 4 genes in the Gilad2004 data were not identified in this study (supplementary table 2, Supplementary Material online). Moreover, 8 genes were regarded as pseudogenes in the Gilad2005 data but were annotated as functional in both the Gilad2004 data and in this study. These observations suggest that the quality of chimpanzee OR genes in this study is greatly improved compared with that in the Gilad2005 data.

Humans have a group of pseudogenes named 7E (Newman and Trask 2003) or H* pseudogenes (Niimura and Nei 2005a). Most of these pseudogenes appear to have been generated by gene duplications after they were pseudogenized (Niimura and Nei 2005a). Thus, they apparently did not contribute to olfactory ability in the ancestral species. For this reason, we examined the fractions of pseudogenes after excluding H* pseudogenes. We identified 84, 85, and 8 H* pseudogenes from humans, chimpanzees, and macaques, respectively. The corrected fractions of pseudogenes without H* pseudogenes are very similar among the 3 species (table 1).

Menashe et al. (2006) developed an algorithm to detect candidates of intact OR pseudogenes based on deviations from a functionally crucial consensus (classifier for olfactory receptor pseudogenes [CORP]). We applied the CORP algorithm to the 387 and 380 functional genes in humans and chimpanzees, respectively. As a result, 162 human genes and 156 chimpanzee genes were identified as candidate intact pseudogenes. The fractions of putative intact pseudogenes are also similar between the 2 species (41.9% and 41.1% for humans and chimpanzees, respectively).

Orthologous Relationships of OR Genes among 3 Primates

By conducting phylogenetic analyses and homology searches, we identified 689 human–chimpanzee and 470

Table 2
Number of Orthologous OR Gene Sets in 3 Primates

		Human–Chimpanzee				
Human [712] ^a	Chimpanzee [724] ^a			Mean ω ^b	Median ω	Nucleotide identity ^c
One	One	647				
Functional	Functional		266		0.937	0.713
Pseudogene	Pseudogene		66	(10) ^d		98.77 ± 0.57%
Pseudogene	Functional		70			
Pseudogene	Pseudogene		246	(1) ^d		
One	Multiple	29				
Multiple	One	12				
Multiple	Multiple	1				
Total		689				
		Human–Macaque				
Human [535] ^a	Macaque [498] ^a			Mean ω	Median ω	Nucleotide identity ^c
One	One	404				
Functional	Functional		157		0.442	0.407
Functional	Pseudogene		78	(7) ^d		95.82 ± 1.04%
Pseudogene	Functional		63			
Pseudogene	Pseudogene		106	(1) ^d		
One	Multiple	21				
Multiple	One	39				
Multiple	Multiple	6				
Total		470				

^a Number of genes included in orthologous gene pairs.^b Nine gene pairs showing $dS = 0$ were eliminated from the analysis. dS , number of synonymous substitutions per synonymous site.^c Nucleotide identity. Mean ± standard deviation.^d Number of truncated genes.

human–macaque orthologous gene sets (table 2). Most of the orthologous relationships are one-to-one (647 out of 689 human–chimpanzee orthologous gene sets and 404 out of 470 human–macaque orthologous gene sets), suggesting that gene duplication was not frequent in the evolution of OR genes in the 3 species. We also identified 443 orthologous gene sets among humans, chimpanzees, and macaques (supplementary table 1, Supplementary Material online). Among them, 371 orthologous relationships are one-to-one-to-one trios. The organization of genomic clusters of OR genes is generally well conserved among the 3 species, though there are several cases in which genomic rearrangements have occurred in the regions of OR gene clusters (supplementary fig. 2, Supplementary Material online). Orthologous relationships were identified for almost all OR gene clusters between humans and chimpanzees (supplementary table 1, Supplementary Material online).

We found 266 one-to-one orthologous pairs of functional OR genes between humans and chimpanzees. For all cases, the orthologous genes are located on syntenic chromosomes between these species. The mean of the nucleotide sequence identities between orthologous genes among the 266 pairs is 98.77% (table 2). This value is much lower than the mean (99.41%) of the nucleotide identities among 13,454 orthologous gene pairs between the 2 species and is exactly the same as that in the whole-genome sequences containing noncoding regions (The Chimpanzee Sequencing and Analysis Consortium 2005). We identified 157 one-to-one functional orthologous gene pairs between humans and macaques. The mean nucleotide identity among them is 95.82%, which is considerably lower than the mean in the entire coding sequences (97.5%) but is higher than that in

the whole-genome sequences (93.54%) (The Rhesus Macaque Genome Sequencing and Analysis Consortium 2007). There are 140 one-to-one-to-one functional orthologous gene trios.

We calculated the ratio of nonsynonymous to synonymous substitutions (ω) for each of the 266 human–chimpanzee orthologous gene pairs and the 157 human–macaque orthologous gene pairs. The mean of the ω values for the human–chimpanzee pairs is 0.937 (median 0.713), which is much larger than that for the human–macaque pairs (0.442, median 407; table 2). The distributions of ω values are significantly different between human–chimpanzee and human–macaque comparisons ($P < 2.2 \times 10^{-16}$, Mann–Whitney U test; supplementary fig. 1, Supplementary Material online). Moreover, the fraction of gene pairs with $\omega > 1$ is much higher for the human–chimpanzee pairs (90/266) than for the human–macaque pairs (2/157). It is known that the mutation rate of CpG dinucleotides is >10 times faster than that of the other sites (Sved and Bird 1990). For this reason, we examined whether or not the presence of CpG dinucleotides distorts the distribution of ω . However, we found no significant correlation between the ω values and the fractions of CpG dinucleotides in OR genes (data not shown).

As mentioned above, it was previously suggested that the pseudogenization rate of OR genes is higher in the human lineage than in the chimpanzee lineage (Gilad et al. 2003, 2004, 2005). As shown in table 2, among the 647 one-to-one orthologous pairs between humans and chimpanzees, 66 gene pairs contain a human functional gene and a chimpanzee pseudogene (or a truncated gene) and 70 gene pairs contain a human pseudogene and

a chimpanzee functional gene. To exclude the possibility of sequencing errors in the chimpanzee genome, we examined the Phred quality score at each disruptive mutation in chimpanzee pseudogenes. However, we did not find any sites showing a Phred score <40 (99.99% base-call accuracy). It therefore appears that pseudogenization rates are similar between the human and chimpanzee lineages. Even when all truncated genes in chimpanzees were assumed to be functional, the null hypothesis of an equal rate of pseudogenization between the 2 lineages could not be rejected ($P = 0.21$, χ^2 test). For 28 and 39 genes out of the 66 human-specific and 70 chimpanzee-specific pseudogenes, respectively, their orthologs were regarded as intact pseudogenes by the CORP algorithm (Menashe et al. 2006; see above).

Wang et al. (2006) reported 80 human-specific non-processed pseudogenes, including 36 OR genes, by conducting whole-genome analyses. Almost all (33) these 36 genes are contained in the 70 human-specific OR pseudogenes identified above. As for the remaining 3 genes, their chimpanzee orthologs were annotated as pseudogenes in our data set.

Though the numbers of functional genes are similar between humans and chimpanzees, we found that the repertoires of OR genes differ considerably between the 2 species. Interestingly, 98 of 387 human functional genes are not orthologous to chimpanzee functional genes, and 94 of 380 functional genes in chimpanzees are nonorthologous to those in humans (fig. 1A). In other words, 25% of the functional OR gene repertoires in humans and chimpanzees are species specific. This observation suggests that the MRCA between the 2 species had a larger number of functional OR genes than humans or chimpanzees, and many genes were lost in a lineage-specific manner (see below). Among the 98 human-specific and 94 chimpanzee-specific genes, 40 and 47 genes were assigned to be intact pseudogenes by the CORP algorithm (Menashe et al. 2006; see above). We also investigated the phylogenetic distribution of these species-specific functional genes. The result showed that some species-specific genes are unevenly distributed in the phylogenetic tree (supplementary fig. 3, Supplementary Material online). For example, clade E in supplementary figure 3 (Supplementary Material online) contains 7 human-specific genes but does not contain any chimpanzee-specific genes.

Evolutionary Trend of Gene Gains and Gene Losses in Primates

To investigate the evolutionary trends of gains and losses of OR genes in primates, we counted the numbers of gene gains (gene duplication) and gene losses (elimination from the genome) that occurred in each lineage by using orthologous relationships (supplementary table 1, Supplementary Material online). From these numbers, we estimated the numbers of OR genes in the ancestral species in the evolution of 3 primate species (fig. 1B). Note that the numbers in this figure include both functional genes and pseudogenes. The result showed that there are more genes in the ancestral species than in each of the 3 current

species, suggesting that gene losses were the major trend in the evolution of OR genes in these species. In the macaque lineage, it was estimated that only 32 gene gains have occurred after the divergence from humans/chimpanzees, whereas 258 genes were deleted from the genome in the same period. The human and chimpanzee lineages also lost many genes; the estimated numbers of gene losses are 2 or 3 times larger than those of gene gains (fig. 1B). The number of gene losses in the human (81) and chimpanzee (88) lineages during ~ 6 Myr is nearly the same as that (82) in the human/chimpanzee lineage before the separation of the 2 species during ~ 24 Myr, suggesting that the rate of gene loss has recently accelerated in the human/chimpanzee lineage.

To examine the evolutionary change in the fraction of pseudogenes, we estimated the numbers of functional OR genes and pseudogenes in the MRCA between humans and chimpanzees. For this purpose, we inferred whether an OR gene in the MRCA was functional or nonfunctional by using the parsimony principle (see Materials and Methods). As figure 1C suggests, the MRCA had a much larger number (510) of functional genes and a significantly lower fraction (40.5%) of pseudogenes than do humans and chimpanzees. We also estimated that among the 510 functional genes in the MRCA, 113 became pseudogenes and 19 were deleted from the genome in the human lineage, whereas 109 were pseudogenized and 23 were eliminated in the chimpanzee lineage. These observations again support the notion of rapid deterioration of OR genes in both human and chimpanzee lineages and similar rates of gene loss between the 2 lineages.

As mentioned above, similar numbers of pseudogenization events occurred in the human and chimpanzee lineages. We then ask whether or not the sets of genes that became pseudogenes are similar between the 2 species. For this purpose, we counted the functional genes in the MRCA that were independently pseudogenized in both the human and chimpanzee lineages (see Materials and Methods). We refer to the pseudogenes derived from such MRCA genes as IPOPs. We found that, among the 113 and 109 genes that were pseudogenized in the human and chimpanzee lineages, respectively, 43 genes are IPOPs. The expected number of IPOPs under the assumption that pseudogenization events occurred randomly in the human and chimpanzee lineages is 24, and the probability of the occurrence of 43 or more IPOPs ($P(i \geq 43)$) is $<10^{-5}$ (see Materials and Methods). Therefore, the same functional gene in the MRCA tends to have become a pseudogene independently in the 2 lineages, although we cannot exclude the possibility that some of these genes were already dysfunctional in the MRCA but had not yet acquired any disruptive mutations. This observation is not surprising because some common environmental factors in humans and chimpanzees should have affected the dispensability of genes. However, if we note that $>60\%$ of the pseudogenization events were lineage specific, the effects of environmental factors on OR gene pseudogenization in these lineages appear not to be large.

One may criticize that the fraction of OR pseudogenes in the MRCA calculated above is underestimated because some of the OR pseudogenes in the MRCA might have

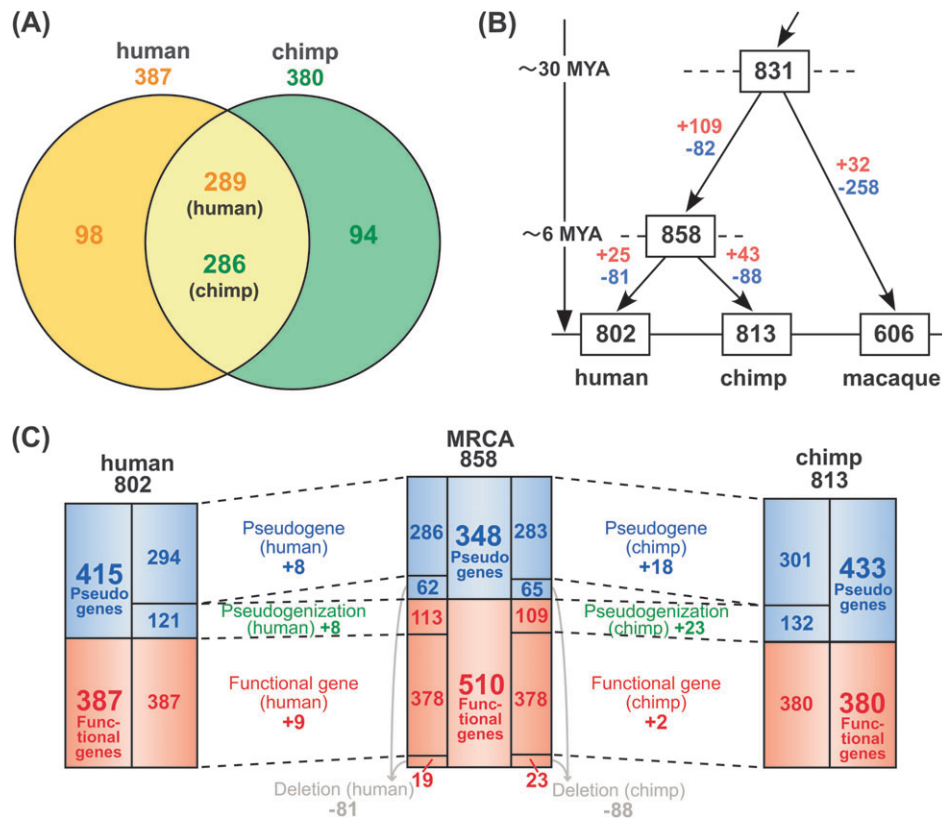


FIG. 1.—The OR gene repertoires in humans and chimpanzees and the numbers of gene gains and losses in primate evolution. (A) Numbers of functional OR genes in humans and chimpanzees. The numbers shown in the overlapped region of 2 circles indicate those of genes originated from common ancestral genes in the MRCA between the 2 species. The numbers for humans and chimpanzees are different because of the different numbers of lineage-specific gene duplications between them. (B) Evolutionary change of the number of OR genes in the genome. A number in a box indicates that of OR genes (including functional genes and pseudogenes) in the extant or ancestral species. Numbers with plus and minus signs indicate those of gene gains and gene losses (elimination from the genome), respectively, in each branch. (C) Numbers of functional genes and pseudogenes in humans, chimpanzees, and their MRCA. The evolutionary changes in the MRCA genes are also illustrated.

been eliminated from both the human and chimpanzee genomes. To consider this possibility, we investigated the genes that were inferred to have existed in the MRCA among humans, chimpanzees, and macaques but that have no surviving descendent genes in both humans and chimpanzees. There are 82 such genes (supplementary table 1, Supplementary Material online). These 82 genes were eliminated from the genome in the human/chimpanzee lineage at some time between 30 MYA and the present. As an extreme case, it is possible to assume that all the 82 genes were present as pseudogenes in the MRCA genome. Even in that case, the fraction of pseudogenes in the MRCA is calculated to be 45.7% ($[348 + 82]/[858 + 82]$), which is still considerably smaller than the fractions in humans and chimpanzees. In reality, some of the 82 genes should have remained functional in the MRCA, and others should already have been eliminated from the MRCA genome. Therefore, our conclusion of the lower fraction of OR pseudogenes in the MRCA than in humans and chimpanzees does not change.

Selective Forces on OR Genes

Several genome-scale studies have suggested that OR genes may have experienced positive selection in either the

human or chimpanzee lineage or both lineages (Clark et al. 2003; Bustamante et al. 2005; Nielsen et al. 2005; Arbiza et al. 2006; Voight et al. 2006). However, some of these studies may be inaccurate because they used distantly related species (e.g., mice or rats) as outgroups (Clark et al. 2003; Arbiza et al. 2006) or did not use any outgroup species (Nielsen et al. 2005). Here we used macaque OR genes as the outgroup and investigated the selective forces to which human and chimpanzee OR genes were subject. We conducted 2 kinds of statistical tests, the branch test (fig. 2A) and the branch-site test (fig. 2B), on the basis of the maximum likelihood method. One-to-one orthologous gene pairs of functional genes between humans and chimpanzees and their orthologous genes in macaques were used for the analyses. We used 219 orthologous gene sets among the 3 species for the following analysis.

In the branch test, we examined 5 different models (Models I–V in fig. 2A). Using PAML (Yang and Nielsen 2000; Yang et al. 2000), we computed likelihood values in each model and chose the best model for each of the orthologous gene sets. We found that 14 and 19 cases supported Models II and III, respectively, at significant levels ($P < 0.05$ in the likelihood ratio test; fig. 2C). Among them, 13 and 16 cases showed acceleration of the evolutionary rates in the human and chimpanzee lineages, respectively.

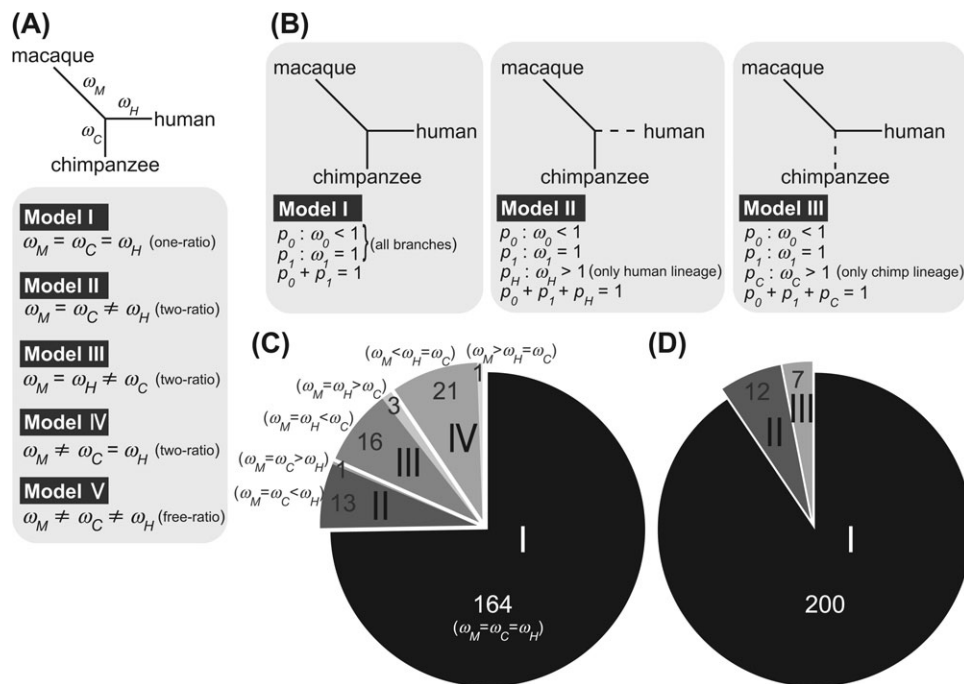


FIG. 2.—Selection test for OR genes in 3 primate species. (A) A schematic phylogenetic tree used for the branch test. The ω values along the human, chimpanzee, and macaque lineages are ω_H , ω_C , and ω_M , respectively. Five different models (Models I–V) were considered. (B) Schematic phylogenetic trees used for the branch-site test. Three models (Models I–III) were assumed. In Models II and III, some codon sites are allowed to be under positive selection in the human and chimpanzee lineages, respectively. (C and D) The numbers of orthologous OR gene sets supporting each model in the branch test (C) and in the branch-site test (D).

The ω values for all these 29 cases are more than 1 (supplementary table 3, Supplementary Material online), possibly suggesting that they were under positive selection. However, it is also possible that these genes are in the process of pseudogenization. We found no significant differences in the number of genes that were under accelerated (13 vs. 16) or decelerated (1 vs. 3) evolution between the human and chimpanzee lineages (fig. 2C). We also corrected the effect of multiple comparisons by controlling the false discovery rate (FDR) (see details in Benjamini and Hochberg 1995), but there were no cases in which a model other than Model I was supported with an FDR level of 0.05.

Because positive selection affects only a subset of codons within a coding region, we carried out the branch-site test, in which the ω value is allowed to vary among codon sites. In this test, we examined 3 models (Models I–III in fig. 2B). In Model I, it is assumed that all sites along all branches are under either negative or neutral selection, whereas in Models II and III, some codons in the human and chimpanzee lineages, respectively, are allowed to evolve under positive selection. The result showed that the numbers of orthologous gene pairs supporting Models II and III are 12 and 7, respectively ($P < 0.05$ in the likelihood ratio test; fig. 2D and supplementary table 4 [Supplementary Material online]). With an FDR level of 0.05, we found only one case supporting Model II and one case supporting Model III. Therefore, there is little difference in the extent of positive selection between the 2 species. Moreover, these results suggest that the contribution of positive

selection to the evolution of OR genes is not large in the human and chimpanzee lineages (Gimelbrant et al. 2004).

Discussion

Comparison with the Previous Studies

In this study, we showed that the fractions of OR pseudogenes are similar between humans and chimpanzees. This result is in sharp contrast to the previous studies (Gilad et al. 2003, 2004, 2005), in which the fraction of OR pseudogenes in chimpanzees was claimed to be significantly lower than that in humans. The discrepancy between the results in this study and those in Gilad et al. (2005) may be explained in the following way. First, Gilad et al. (2005) estimated the fraction of OR pseudogenes in chimpanzees on the basis of only 30 intact genes in Gilad et al. (2003). As mentioned above, the Gilad2005 data included 899 chimpanzee OR gene candidates containing 353 intact genes and 546 pseudogenes. (Note that the 353 intact genes contain many truncated genes.) Among these pseudogenes, they identified 132 7E pseudogenes. By excluding the 7E pseudogenes, they calculated the fraction of pseudogenes in chimpanzees to be $\sim 54\%$. They then compared the Gilad2005 data with the 30 intact OR genes that were experimentally determined in Gilad et al. (2003). They found that 7 of the 30 intact genes were regarded as pseudogenes in the Gilad2005 data and assumed that this discrepancy was due to sequencing errors in the genome sequences used. Using these numbers, they calculated the corrected fraction of OR

pseudogenes in chimpanzees to be ~41%. We should note that this estimation would contain a large extent of sampling errors because it is based on only 30 genes.

The second reason for the discrepancy is the difference in the fraction of pseudogenes in humans. Gilad et al. (2005) mentioned that, on the basis of the HORDE (<http://biportal.weizmann.ac.il/HORDE/>), the fraction of OR pseudogenes in humans after excluding 7E pseudogenes was ~51%. This value is considerably higher than our result (table 1). However, according to the latest version of HORDE (Build 42), which contains 391 functional genes and 464 pseudogenes including 85 7E pseudogenes in humans, the fraction of OR pseudogenes without 7E pseudogenes is calculated to be ~49%. This value is closer to, but still higher than, that in table 1. The smaller number of OR pseudogenes in our data than in HORDE might be due to the slightly conservative criteria for our OR gene identification (Niimura and Nei 2007), suggesting that the fraction of pseudogenes can vary depending on the gene identification method. The finding of similar fractions of pseudogenes between humans and chimpanzees was obtained by applying the same criteria to the genome sequences of the 2 species. Therefore, the result in this paper is expected to be more reliable than that in Gilad et al. (2005).

In another paper, Gilad et al. (2004) also claimed that the fraction of OR pseudogenes is significantly higher in humans than in chimpanzees, based on a survey of 100 randomly chosen OR genes. There are a number of mistakes in that paper (see Gilad et al. 2007). After the correction of these mistakes, there are 38 pseudogenes in chimpanzees in the Gilad2004 data (supplementary table 2, Supplementary Material online). We should note, however, that the fraction obtained from these data (38%) does not represent the fraction of OR pseudogenes in the chimpanzee genome. The reason is that they used degenerate primers that covered only ~670 bp out of a ~1-kb coding region of an OR gene. Therefore, pseudogenes that carry disruptive nonsense or frameshift mutations only in the region not amplified by these primers were regarded to be intact in the Gilad2004 data. In fact, we found 6 such cases (*PC1_17_i*, *PC2_29_i*, *PC2_55_i*, *PC2_56_i*, *PC2_57_i*, and *PC2_69_i*; see supplementary table 2, Supplementary Material online). If we use this number, the fraction of pseudogenes without 7E genes is estimated to be 44%, which is fairly close to the result in table 1.

Evolution of Human and Chimpanzee OR Genes

This paper revealed that, although the numbers of functional genes and pseudogenes are similar between humans and chimpanzees, ~25% of their functional gene repertoires are nonorthologous to each other (fig. 1A). The difference in the functional OR gene repertoires between the 2 species might be responsible for the difference in the ability of odorant detection and may affect their behaviors. Niimura and Nei (2007) showed that, although the numbers of functional OR genes are similar (~1,000) among several mammalian species (mice, rats, dogs, cows, and opossums), their OR gene repertoires were generated after the occurrence of hundreds of gene gains and losses

from their common ancestor. Therefore, it appears that OR gene repertoires in mammals are generally highly diverse among different species.

Olfactory information is first transmitted to the olfactory bulb, which is involved in the integration of signals from different ORs (Mori et al. 1999). It is therefore reasonable to assume that the size of the olfactory bulb is a good indicator of olfactory ability. Comparative anatomical studies showed that the olfactory bulb in chimpanzees is more than twice as large as that in humans (114 mm³ for humans vs. 257 mm³ for chimpanzees) (Stephan et al. 1981). The difference becomes even more prominent when the size of the olfactory bulb is normalized by body weight and brain volume (Schoenemann 1997). Moreover, it was reported that wild chimpanzees use several olfactory cues in both social and sexual contexts (Goodall 1986; Nishida 1997; Boesch and Boesch-Achermann 2000). These observations may suggest that the olfactory ability is better in chimpanzees than in humans. However, to our knowledge, there have been no quantitative studies comparing the olfactory abilities between humans and chimpanzees. It is also possible that the number of functional genes is not necessarily well correlated with olfactory ability in each species. For example, dogs are generally regarded to have a keen sense of olfaction, but they do not have a particularly large number of functional OR genes (~800) (Niimura and Nei 2007).

We should note that, however, the exact numbers of functional genes and pseudogenes in humans and chimpanzees cannot be determined in practice. Recently Nozawa et al. (2007) investigated the copy number variation of OR genes among 270 humans. They found that the difference in the number of functional OR genes between 2 randomly chosen individuals is ~11 on average. Moreover, the extent of copy number variation is almost the same between functional genes and pseudogenes, suggesting that a small change of the number of OR genes is nearly neutral. It is also known that there are many OR segregating pseudogenes that display both functional and nonfunctional alleles in humans (Menashe et al. 2003, 2007).

The results in figure 1B and C clearly showed that the OR gene repertoires in the 3 primate species are in a phase of deterioration. These species may have lost OR genes because they possess trichromatic vision, as previously suggested (Gilad et al. 2004). Although a general evolutionary trend of OR genes in primates is the reduction of functional constraints, it is possible that some genes are under positive selection (Clark et al. 2003; Bustamante et al. 2005; Nielsen et al. 2005; Arbiza et al. 2006; Voight et al. 2006). We therefore examined whether or not OR genes were subject to positive selection by using PAML (Yang and Nielsen 2000; Yang et al. 2000). PAML is often used to detect lineage-specific or site-specific positive selection (Clark et al. 2003; Gilad et al. 2005; Arbiza et al. 2006; Yu et al. 2006), but this method is known to produce false positives under certain conditions (Suzuki and Nei 2002, 2004). However, our goal is not to count the number of genes that are subject to positive selection but to compare evolutionary forces between the human and chimpanzee lineages. Biases introduced by PAML should affect human and chimpanzee genes equally. As a result, we found no

significant difference in the number of genes that were under accelerated evolution (fig. 2C and D). Because it is known that a closely related species used as an outgroup gives a more accurate estimation of lineage-specific changes (Yu et al. 2006; Bakewell et al. 2007), our use of macaques rather than mice as the outgroup is expected to improve the accuracy of the results.

By the branch-site test with an FDR level of 0.05, we found only one case (*HsOR7.6.15*) in which positive selection was suggested in the human lineage. This result is consistent with Gimelbrant et al. (2004); that study suggested that the effect of positive selection on the evolution of OR genes in the human lineage is not large. In the *HsOR7.6.15* gene, 13 nonsynonymous substitutions and only 1 synonymous substitution were inferred to have occurred along the human lineage. On the other hand, in the chimpanzee lineage leading to the chimpanzee ortholog (*PatrOR7.6.18*), it was inferred that 4 nonsynonymous and 5 synonymous changes occurred. Therefore, *HsOR7.6.15* is a candidate gene that has experienced human-specific positive selection. Among the 13 nonsynonymous substitutions that have taken place in the human lineage, 1 substitution (from methionine to threonine at position 259) occurred at a putative ligand-binding site (Katada et al. 2005), which might have altered ligand-binding affinity in a way that was advantageous to humans. However, further experimental studies should be conducted in order to elucidate how the difference in OR gene repertoires between the 2 species is responsible for the difference in physiological and behavioral traits.

Supplementary Material

Supplementary data sets 1 and 2, figures 1–3, and tables 1–4 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We thank Pierre Fontanillas, Erik Dopman, Rob Kulathinal, and Trevor Bedford for their critical readings of the manuscript. We thank Doron Lancet and Idan Menashe for providing us the perl script of the CORP algorithm and Masafumi Nozawa for the perl script to calculate ω values. We also thank Masatoshi Nei and Daniel Hartl for stimulating discussions and comments on the manuscript. This study was supported by the Japanese Society for Promotion of Science grant 17-02667 (Y.G.), 19-589 (Y.G.), and by the Ministry of Education, Culture, Sports, Science and Technology, Japan, grant 17710162 and 20770192 (Y.N.).

Literature Cited

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped Blast and PSI-Blast: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
- Arbiza L, Dopazo J, Dopazo H. 2006. Positive selection, relaxation, and acceleration in the evolution of the human and chimpanzee genome. *PLoS Comput Biol.* 2:e38.
- Bakewell MA, Shi P, Zhang J. 2007. More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proc Natl Acad Sci USA.* 104:7489–7494.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B.* 57:289–300.
- Boesch C, Boesch-Achermann H. 2000. *The chimpanzees of the Tai forest: behavioural ecology and evolution.* Oxford: Oxford University Press.
- Buck L, Axel R. 1991. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell.* 65:175–187.
- Bustamante CD, Fledel-Alon A, Williamson S, et al. (14 co-authors). 2005. Natural selection on protein-coding genes in the human genome. *Nature.* 20:1153–1157.
- Clark AG, Glanowski S, Nielsen R, et al. (17 co-authors). 2003. Inferring nonneutral evolution from human-chimpanzee-mouse orthologous gene trios. *Science.* 302:1960–1963.
- Gilad Y, Man O, Glusman G. 2005. A comparison of the human and chimpanzee olfactory receptor gene repertoires. *Genome Res.* 15:224–230.
- Gilad Y, Man O, Pääbo S, Lancet D. 2003. Human specific loss of olfactory receptor genes. *Proc Natl Acad Sci USA.* 100:3324–3327.
- Gilad Y, Wiebe V, Przeworski M, Lancet D, Pääbo S. 2004. Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol.* 2:e5.
- Gilad Y, Wiebe V, Przeworski M, Lancet D, Pääbo S. 2007. Correction: loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol.* 5:e148.
- Gimelbrant AA, Skaletsky H, Chess A. 2004. Selective pressures on the olfactory receptor repertoire since the human-chimpanzee divergence. *Proc Natl Acad Sci USA.* 101:9019–9022.
- 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 PA, Malnic B, Buck LB. 2004. The mouse olfactory receptor gene family. *Proc Natl Acad Sci USA.* 101:2156–2161.
- Goodall J. 1986. *The chimpanzees of Gombe.* Cambridge (MA): Harvard University Press.
- Katada S, Hirokawa T, Oka Y, Suwa M, Touhara K. 2005. Structural basis for a broad but selective ligand spectrum of a mouse olfactory receptor: mapping the odorant-binding site. *J Neurosci.* 25:1806–1815.
- Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform.* 5:150–163.
- Menashe I, Abaffy T, Hasin Y, Goshen S, Yahalom V, Luetje CW, Lancet D. 2007. Genetic elucidation of human hyperosmia to isovaleric acid. *PLoS Biol.* 5:e284.
- Menashe I, Aloni R, Lancet D. 2006. A probabilistic classifier for olfactory receptor pseudogenes. *BMC Bioinformatics.* 7:393.
- Menashe I, Man O, Lancet D, Gilad Y. 2003. Different noses for different people. *Nat Genet.* 34:143–44.
- Mombaerts P. 2004. Genes and ligands for odorant, vomeronasal and taste receptors. *Nat Rev Neurosci.* 5:263–278.
- Mori K, Nagao H, Yoshihara Y. 1999. The olfactory bulb: coding and processing of odor molecule information. *Science.* 286:711–715.
- Newman T, Trask BJ. 2003. Complex evolution of 7E olfactory receptor genes in segmental duplications. *Genome Res.* 13:781–793.

- Nielsen R, Bustamante C, Clark AG, et al. (13 co-authors). 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* 3:e170.
- Niimura Y, Nei M. 2003. Evolution of olfactory receptor genes in the human genome. *Proc Natl Acad Sci USA.* 100:12235–12240.
- Niimura Y, Nei M. 2005a. Comparative evolutionary analysis of olfactory receptor gene clusters between humans and mice. *Gene.* 346:13–21.
- Niimura Y, Nei M. 2005b. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proc Natl Acad Sci USA.* 102:6039–6044.
- Niimura Y, Nei M. 2006. Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates. *J Hum Genet.* 51:505–517.
- Niimura Y, Nei M. 2007. Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE.* 2:e708.
- Nishida T. 1997. Sexual behavior of adult male chimpanzees of the Mahale Mountains National Park, Tanzania. *Primates.* 38:379–398.
- Nozawa M, Kawahara Y, Nei M. 2007. Genomic drift and copy number variation of sensory receptor genes in humans. *Proc Natl Acad Sci USA.* 104:20421–20426.
- Olander T, Fuchs T, Linhart C, Shamir R, Adams M, Kalush F, Khen M, Lancet D. 2004. The canine olfactory subgenome. *Genomics.* 83:361–372.
- Quignon P, Giraud M, Rimbault M, et al. (11 co-authors). 2005. The dog and rat olfactory receptor repertoires. *Genome Biol.* 6:R83.
- Quignon P, Kirkness E, Cadiou E, Touleimat N, Guyon R, Renier C, Hitte C, Andre C, Fraser C, Galibert F. 2003. Comparison of the canine and human olfactory receptor gene repertoires. *Genome Biol.* 4:R80.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 4:406–425.
- Schoenemann PT. 1997. An MRI study of the relationship between human neuroanatomy and behavioral ability [dissertation]. [Berkeley (CA)]: University of California.
- Stephan H, Frahm H, Baron G. 1981. New and revised data on volumes of brain structures in insectivores and primates. *Folia Primatol.* 35:1–29.
- Suzuki Y, Nei M. 2002. Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. *Mol Biol Evol.* 19:1865–1869.
- Suzuki Y, Nei M. 2004. False-positive selection identified by ML-based methods: examples from the *Sigl* gene of the diatom *Thalassiosira weissflogii* and the *tax* gene of a human T-cell lymphotropic virus. *Mol Biol Evol.* 21:914–921.
- Sved J, Bird A. 1990. The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. *Proc Natl Acad Sci USA.* 87:4692–4696.
- The Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature.* 437:69–87.
- The Rhesus Macaque Genome Sequencing and Analysis Consortium. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science.* 316:222–234.
- Thompson JD, Higgins DG, Gibson TJ. 1994. ClustalW: 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.
- Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72.
- Wang X, Grus W, Zhang J. 2006. Gene losses during human origins. *PLoS Biol.* 4:e52.
- Yang Z, Nielsen R. 2000. Estimating synonymous and non-synonymous substitution rates under realistic evolutionary models. *Mol Biol Evol.* 17:32–43.
- Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics.* 155:431–449.
- Young JM, Friedman C, Williams EM, Ross JA, Tonnes-Priddy L, Trask BJ. 2002. Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Hum Mol Genet.* 11:535–546.
- Yu XJ, Zheng HK, Wang J, Wang W, Su B. 2006. Detecting lineage-specific adaptive evolution of brain-expressed genes in human using rhesus macaque as outgroup. *Genomics.* 88:745–751.
- Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol.* 22:2472–2479.
- Zhang J, Rosenberg HF, Nei M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci USA.* 95:3708–3713.
- 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–research0018.12.

Adriana Briscoe, Associate Editor

Accepted June 10, 2008