

# Lineage-Specific Loss of Function of Bitter Taste Receptor Genes in Humans and Nonhuman Primates

Yasuhiro Go,<sup>\*,1</sup> Yoko Satta,<sup>\*</sup> Osamu Takenaka<sup>†,2</sup> and Naoyuki Takahata<sup>\*</sup>

<sup>\*</sup>Department of Biosystems Science, Graduate University for Advanced Studies (Sokendai), Hayama, Kanagawa 240-0193, Japan and <sup>†</sup>Primate Research Institute, Kyoto University, Inuyama, Aichi 484-8506, Japan

Manuscript received October 15, 2004  
Accepted for publication January 28, 2005

## ABSTRACT

Since the process of becoming dead genes or pseudogenes (pseudogenization) is irreversible and can occur rather rapidly under certain environmental circumstances, it is one plausible determinant for characterizing species specificity. To test this evolutionary hypothesis, we analyzed the tempo and mode of duplication and pseudogenization of bitter taste receptor (*T2R*) genes in humans as well as in 12 nonhuman primates. The results show that primates have accumulated more pseudogenes than mice after their separation from the common ancestor and that lineage-specific pseudogenization becomes more conspicuous in humans than in nonhuman primates. Although positive selection has operated on some amino acids in extracellular domains, functional constraints against *T2R* genes are more relaxed in primates than in mice and this trend has culminated in the rapid deterioration of the bitter-tasting capability in humans. Since *T2R* molecules play an important role in avoiding generally bitter toxic and harmful substances, substantial modification of the *T2R* gene repertoire is likely to reflect different responses to changes in the environment and to result from species-specific food preference during primate evolution.

**N**O doubt, organisms have increasingly acquired new genes by tandem duplication and/or polyploidization (OHNO 1970). However, as evolution has proceeded, loss of genes has also occurred whenever they became dispensable under certain circumstances of interplay between genes and their environments. In animals, nutrition that is ingested daily from surroundings must have been one of the most important environmental factors and must have permitted genes involved in vitamin C biosynthesis (NISHIKIMI *et al.* 1994) and essential amino acid biosynthesis (LEHNINGER 1996) to be non-functional. A similar phenomenon may be found in genes for the sense of taste since tasting plays a crucial role in providing animals with information about proper assessments of foods and appropriate behaviors. Mammals can basically sense tastes of sweet, sour, bitterness, salt, and umami (the taste of monosodium glutamate). Of these five modalities, sweet, bitter, and umami substances are perceived by G-protein-coupled receptor (GPCR) signaling pathways (WONG *et al.* 1996). Two GPCR families are involved in these pathways: One is *T1R*, which is associated with sweet and umami substances (NELSON *et al.* 2001; LI *et al.* 2002), and the other is *T2R*, which is associated with bitter substances

(ADLER *et al.* 2000; CHANDRASHEKAR *et al.* 2000; MATSUNAMI *et al.* 2000). Unfortunately, however, the correspondence between bitter substances (ligands) and receptors is poorly known. One exception is one of the human *T2R* genes, which is proven to be responsible for perception of phenylthiocarbamide (KIM *et al.* 2003).

Both *T1R* and *T2R* have seven transmembrane domains as a common character of GPCRs. Despite this structural similarity, these two molecules exhibit no obvious amino acid sequence similarities. Rather, *T1R* and *T2R* are closely related to the pheromone receptors *V2R* and *V1R*, respectively (MATSUNAMI and AMREIN 2003). *T2R* genes form a larger multigene family than *T1R* genes. In contrast to the presence of only 3 *T1R* genes in the mammalian genome (NELSON *et al.* 2001; LI *et al.* 2002), >30 *T2R* genes exist (CONTE *et al.* 2003; SHI *et al.* 2003). The recent chicken genome project (HILLIER *et al.* 2004) ascertained that the *T2R* gene repertoire had expanded in the ancestral mammalian lineage after its divergence from reptiles/birds and that the capacity for sensing bitter substances has broadened in mammals relative to chickens, which possess only 3 *T2R* genes. Between mammalian orders such as humans and mice, the orthologous relationships of *T2R* genes are found to be either one to one or one to multiple (CONTE *et al.* 2003; SHI *et al.* 2003). Since the one-to-multiple orthology has resulted from lineage-specific gene duplication, the *T2R* gene family has continuously undergone gene duplication even after the mammalian radiation. It is also known that the human and mouse *T2R* gene families contain eight and three pseudogenes, re-

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AB198983–AB199308.

<sup>1</sup>Corresponding author: Department of Biosystems Science, Graduate University for Advanced Studies (Sokendai), Hayama, Kanagawa 240-0193, Japan. E-mail: go\_yasuhiro@soken.ac.jp

<sup>2</sup>Deceased.

spectively (CONTE *et al.* 2003; SHI *et al.* 2003). The absence of orthologous pseudogenes between humans and mice immediately indicates that pseudogenization [the process of becoming pseudogenes by disrupting intact open reading frames (ORFs)] of *T2R* genes occurred independently in mammalian lineages. Since pseudogenization can occur rather rapidly under certain circumstances, we can regard it as one possible determinant for characterizing species specificity. This view was actually exemplified by the olfactory receptor (*OR*) gene (GILAD *et al.* 2003, 2004) and pheromone receptor (*VR*) gene families (ZHANG and WEBB 2003). In these families, more pseudogenes have accumulated in hominoids and Old World monkeys (OWMs) than in other mammals and it has been argued that the transition from smell or pheromone dependency to full trichromatic color vision dependency in their life might have relaxed functional constraints against *OR* and *VR* genes (GILAD *et al.* 2003, 2004; ZHANG and WEBB 2003).

Because of important roles of *T2R* in avoiding generally bitter, toxic, and harmful substances, changes in the *T2R* gene repertoire could in turn affect the behavior of animals, especially the feeding behavior. Here, to gain insights into evolutionary changes in the *T2R* gene repertoire, we have conducted a comparative analysis of *T2R* genes among humans, nonhuman primates, tupais, and mice. We show that lineage-specific gene duplication and pseudogenization played important roles in altering the *T2R* gene repertoire in individual primates.

## MATERIALS AND METHODS

**Sources of DNA samples:** Genomic DNAs used here were for the chimpanzee (*Pan troglodytes*; Patr), the gorilla (*Gorilla gorilla*; Gogo), the orangutan (*Pongo pygmaeus*; Popy), the agile gibbon (*Hyllobates agilis*; Hyag), the rhesus monkey (*Macaca mulatta*; Mamu), the silvered leaf monkey (*Trachypithecus cristatus*; Trcr), the white-tufted-ear marmoset (*Callithrix jacchus*; Caja), the brown capuchin (*Cebus apella*; Ceap), the thick-tailed bush baby (*Otolemur crassicaudatus*; Otrc), the Senegal galago (*Galago senegalensis*; Gase), the slow loris (*Nycticebus coucang*; Nyco), the ring-tailed lemur (*Lemur catta*; Leca), and the common tree shrew (*Tupaia glis*; Tugl). These DNAs were generously provided by Primate Research Institute, Kyoto University. Genomic DNAs for human (*Homo sapiens*; Hosa) populations were purchased from Coriell (Camden, NJ).

**PCR amplification and sequencing:** Genomic PCR was performed with the polymerase KOD-Plus under conditions recommended by the manufacturer (TOYOBO). Because all *T2R* genes are intronless and ~900 bp in size, the nucleotide sequences of the entire coding region were obtained by single PCR. The first primer sets for PCR were designed for each *T2R* coding region to amplify ORFs as long as possible. For some closely related *T2R* genes, degenerated primers were designed. In case of failure of amplification by the first primer sets, other primer sets were designed for conserved regions between human and mouse *T2R* homologous genes. The primer sequences are available in supplemental Table 1 at <http://www.genetics.org/supplemental/>. Amplifications were carried out under the following standard conditions: denaturation at 94° for 3 min; 35 cycles of 94° for 20 sec, 45°–52° for

30 sec, and 68° for 30 sec with an additional extension at 68° for 5 min. PCR products were electrophoresed using 1.5% agarose gel and purified by the QIAquick gel extraction kit (QIAGEN, Chatsworth, CA). PCR products were directly sequenced or sequenced after cloning (Zero Blunt TOPO PCR cloning kit, QIAGEN).

**Data analysis:** Sequences were aligned by CLUSTAL\_X (THOMPSON *et al.* 1997) and the resulting alignment was manually adjusted. Codon positions specified in this study correspond to those in the aligned *T2R* sequences, in which the numbering starts at the first codon of *Hosa-T2R39*. Phylogenetic trees were reconstructed by the neighbor-joining (NJ) (SAITOU and NEI 1987) and/or maximum parsimony (FITCH 1971; HARTIGAN 1973) methods implemented in MEGA2 (KUMAR *et al.* 2001). In the NJ trees, the *p* distances for nucleotide or amino acid sequences were used to infer the topology (SAITOU and NEI 1987). To examine whether natural selection has been exerted on individual amino acids, the computer program of DataMonkey (<http://www.datamonkey.org/>) was used (KOSAKOVSKY POND and FROST 2005). This program involves five steps. First, to obtain a reliable estimate of the number of substitutions occurred at a single amino acid site, a nucleotide substitution model fitted to the data was selected by the likelihood ratio test and/or Akaike information criteria score. Second, to obtain ancestral sequences, a codon-based substitution model fitted to the data was selected using the maximum-likelihood method with the MG94 model (MUSE and GAUT 1994). Third, the average numbers of synonymous ( $E_S$ ) and nonsynonymous ( $E_N$ ) sites in a particular phylogenetic tree for each codon were computed, as in SUZUKI and GOJOBORI (1999), by considering the nucleotide substitution biases estimated from the data. Fourth, the total numbers of synonymous ( $N_S$ ) and nonsynonymous ( $N_N$ ) substitutions throughout the phylogenetic tree were counted for each codon site. Finally, we tested whether the number of nonsynonymous substitutions per nonsynonymous site ( $d_N$ ;  $N_N/E_N$ ) is significantly different from the number of synonymous substitutions per synonymous site ( $d_S$ ;  $N_S/E_S$ ) using the extended binomial distribution. In addition, to estimate the numbers of per-site synonymous substitutions ( $b_S$ ) and per-site nonsynonymous substitutions ( $b_N$ ) for each branch of a given tree, the computer program Bn-Bs (ZHANG *et al.* 1998) was used.

## RESULTS

**Comparative analysis of *T2R* genes between humans and mice:** Previous studies identified 33~34 *T2R* genes in humans and 36~40 in mice (CONTE *et al.* 2002, 2003; SHI *et al.* 2003). We reexamined *T2R* genes for the human (NCBI; Build 34 Version 2) and the mouse (NCBI; Build 32 Version 1) genome sequences by BLASTN and TBLASTN with all annotated human and mouse *T2R* genes as queries. These database searches identified 1 additional mouse *T2R* pseudogene (*mt2r37\**; an asterisk indicates a pseudogene), so that there are 41 paralogous genes in mice, including 6 pseudogenes. In humans, on the other hand, 25 *T2R* functional genes and 11 pseudogenes were confirmed. Among 11 pseudogenes, 6 were already identified in both CONTE *et al.* (2002) and SHI *et al.* (2003) and the remaining 5 by one of these studies. The proportion of *T2R* pseudogenes is 31% (11/36) in humans and 15% (6/41) in mice (Table 1).

The NJ tree was reconstructed on the basis of amino acid sequences of all human and mouse *T2R* genes

TABLE 1  
Nomenclature of the mouse *t2r* genes and their human orthologous genes

Mouse <i>t2r</i> genes		SHI <i>et al.</i> (2003)	CONTE <i>et al.</i> (2003)	Category	Human <i>T2R</i> genes		SHI <i>et al.</i> (2003)	CONTE <i>et al.</i> (2002)
This study	Chromosome				This study	Chromosome		
<i>mt2r1</i>	6		<i>mTAS2R33</i>	C	<u><i>hT2R2</i></u> (*)	7	<i>hTAS2R2</i>	
<i>mt2r2</i>	6	<i>mt2r40</i>	<i>mTAS2R18</i>	C	<i>hT2R16</i>	7	<i>ht2r16</i>	
<i>mt2r3</i>	6	<i>mt2r41</i>	<i>mTAS2R37</i>	C	<i>hT2R3</i>	7	<i>ht2r3</i>	
<i>mt2r4</i>	6	<i>mt2r8</i>	<i>mTAS2R8</i>	C	<i>hT2R4</i>	7	<i>ht2r4</i>	
					<u><i>hT2R66</i></u> *	7	<u><i>hps3</i></u>	
					<u><i>hT2R5</i></u>	7	<i>ht2r5</i>	
<i>mt2r5</i>	6	<i>mt2r31</i>	<i>mTAS2R38</i>	C	<i>hT2R38</i>	7	<i>ht2r38</i>	
<i>mt2r6</i>	6	<i>mt2r34</i>	<i>mTAS2R39</i>	C	<i>hT2R39</i>	7	<i>ht2r39</i>	
<i>mt2r7</i>	6	<i>mt2r33</i>	<i>mTAS2R44</i>	C	<i>hT2R40</i>	7	<i>ht2r40</i>	
<i>mt2r8</i>	6	<i>mt2r36</i>	<i>mTAS2R43</i>	C	<u><i>hT2R62</i></u> *	7	<u><i>hps1</i></u>	
<i>mt2r9</i>	6	<i>mt2r38</i>	<i>mTAS2R35</i>	C	<u><i>hT2R56</i></u>	7	<i>ht2r56</i>	
<i>mt2r10</i>	6	<i>mt2r35</i>	<i>mTAS2R26</i>	C	<i>hT2R41</i>	7	<i>ht2r41</i>	
<i>mt2r11</i>	6	<i>mt2r42</i>	<i>mTAS2R30</i>	C	<i>hT2R7</i>	12	<i>ht2r7</i>	
					<i>hT2R8</i>	12	<i>ht2r8</i>	
					<i>hT2R9</i>	12	<i>ht2r9</i>	
<i>mt2r12</i>	6	<i>mt2r43</i>	<i>mTAS2R7</i>	B				
<i>mt2r13</i>	6	<i>mt2r44</i>	<i>mTAS2R6</i>	B				
<i>mt2r14</i>	6	<i>mt2r45</i>	<i>mTAS2R4</i>	B	<i>hT2R10</i>	12	<i>ht2r10</i>	
<i>mt2r15</i>	6	<i>mt2r5</i>	<i>mTAS2R5</i>	B				
<i>mt2r16</i>	6	<i>mt2r46</i>	<i>mTAS2R14</i>	B				
				A	<u><i>hT2R15</i></u> *	12	<u><i>hps8</i></u>	
				A	<i>hT2R50</i>	12	<i>ht2r50</i>	
				A	<i>hT2R49</i>	12	<i>ht2r49</i>	
				A	<i>hT2R48</i>	12	<i>ht2r48</i>	
				A	<u><i>hT2R64</i></u> *	12	<u><i>hps2</i></u>	
<i>mt2r17</i>	6	<i>mt2r47</i>	<i>mTAS2R20</i>	A	<u><i>hT2R63</i></u> *	12	<u><i>hps6</i></u>	
<i>mt2r23</i>	6	<i>mt2r52</i>	<i>mTAS2R36</i>	A	<u><i>hT2R68</i></u> *	12	<u><i>hps7</i></u>	
				A	<u><i>hT2R43</i></u>	12	<i>ht2r43</i>	
				A	<i>hT2R44</i>	12	<i>ht2r44</i>	
				A	<i>hT2R45</i>	12	<i>ht2r45</i>	
				A	<i>hT2R46</i>	12	<i>ht2r46</i>	
				A	<i>hT2R47</i>	12	<i>ht2r47</i>	
				A	<u><i>hT2R11</i></u> *	12	<u><i>hTAS2R11</i></u>	
<i>mt2r18</i>	6	<i>mt2r48</i>	<i>mTAS2R21</i>	B				
<i>mt2r21</i>	6	<i>mt2r50</i>	<i>mTAS2R24</i>	B	<i>hT2R13</i>	12	<i>ht2r13</i>	
<i>mt2r22</i>	6	<i>mt2r51</i>	<i>mTAS2R2</i>	B				
<i>mt2r19</i>	6		<i>mTAS2R22</i>	C	<u><i>hT2R12</i></u> *	12	<u><i>hTAS2R12</i></u>	
<i>mt2r20</i>	6	<i>mt2r49</i>	<i>mTAS2R15</i>	B				
<u><i>mt2r24</i></u> *	6	<u><i>mps3</i></u>	<u><i>mTAS2R45</i></u>	B				
<i>mt2r25</i>	6	<i>mt2r54</i>	<i>mTAS2R17</i>	B				
<i>mt2r26</i>	6	<i>mt2r55</i>	<i>mTAS2R23</i>	B				
<i>mt2r27</i>	6	<i>mt2r56</i>	<i>mTAS2R16</i>	B				
<i>mt2r28</i>	6	<i>mt2r57</i>	<i>mTAS2R10</i>	B				
<i>mt2r29</i>	6	<i>mt2r58</i>	<i>mTAS2R13</i>	B				
<u><i>mt2r30</i></u> *	6		<u><i>mTAS2R42</i></u>	B				
<i>mt2r31</i>	6	<i>mt2r59</i>	<i>mTAS2R25</i>	B	<i>hT2R14</i>	12	<i>ht2r14</i>	
<u><i>mt2r32</i></u> *	6	<u><i>mps1</i></u>	<u><i>mTAS2R46</i></u>	B				
<i>mt2r33</i>	6	<i>mt2r60</i>	<i>mTAS2R29</i>	B				
<i>mt2r35</i>	6	<i>mt2r62</i>	<i>mTAS2R9</i>	B				
<u><i>mt2r36</i></u> *	6	<u><i>mps2</i></u>	<u><i>mTAS2R11</i></u>	B				
<u><i>mt2r37</i></u> *	6			B				
<i>mt2r38</i>	6	<i>mt2r63</i>	<i>mTAS2R3</i>	B				
<i>mt2r39</i>	6	<i>mt2r64</i>	<i>mTAS2R40</i>	B				
<u><i>mt2r40</i></u> *	6		<u><i>mTAS2R41</i></u>	B				
				A	<u><i>hT2R65</i></u> *	12	<u><i>hps4</i></u>	
<i>mt2r34</i>	6	<i>mt2r61</i>	<i>mTAS2R31</i>	A	<u><i>hT2R67</i></u> *	12	<u><i>hps5</i></u>	
				A	<u><i>hT2R55</i></u>	12	<i>ht2r55</i>	
<i>mt2r41</i>	15	<i>mt2r19</i>	<i>mTAS2R19</i>	C	<i>hT2R1</i>	5	<i>ht2r1</i>	

Pseudogenes are underlined.

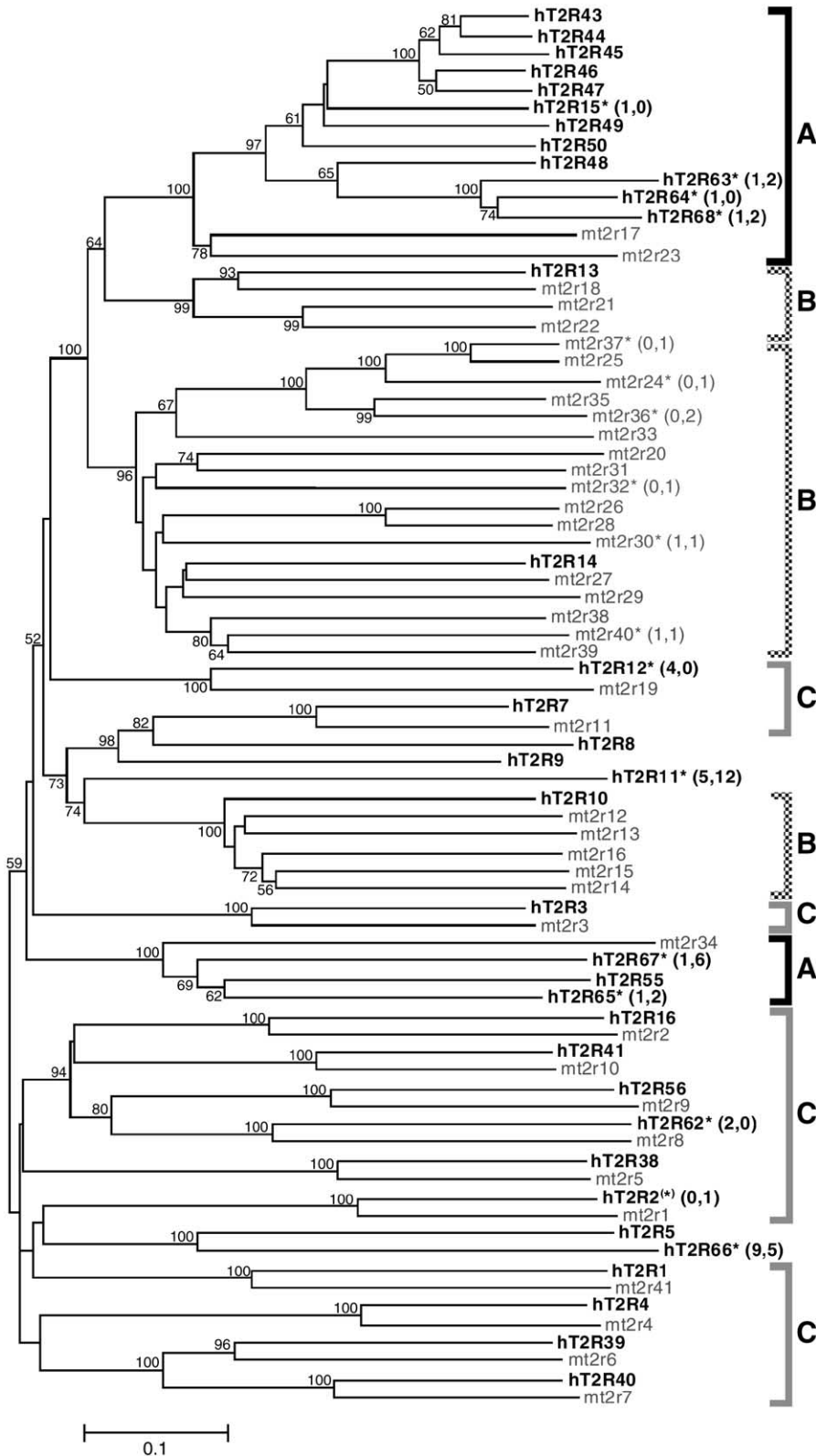


FIGURE 1.—Phylogenetic tree of bitter taste receptor (*T2R*) genes based on amino acid sequences. Genes in humans and mice are shown as solid and shaded, respectively. The tree was reconstructed by the NJ method (SAITOU and NEI 1987) on the basis of the *p* distances, and only >50% bootstrap values are shown at each node (1000 replications). Pseudogenes in this study are defined by the presence of ORF disruptive mutations and are indicated by asterisks (\*). The codon frames in pseudogenes are inferred from those of closely related functional genes. The nomenclature of human *T2R* genes is the same as that in CONTE *et al.* (2002) and SHI *et al.* (2003). However, because of the different designation systems of the mouse genes used by two research groups (CONTE *et al.* 2002, 2003; SHI *et al.* 2003), we have re-named the mouse genes. Detailed nomenclatures in humans and mice are in Table 1. The numbers in parentheses stand for those of nonsense and indel mutations, respectively. The definitions of categories A, B, and C are given in the text.

(Figure 1) and the evolutionary relationships between human and mouse *T2R* genes can be divided into three categories with respect to their orthology (Table 1). In category A, a group of three human genes is orthologous to one mouse gene (multiple-to-one orthology). In addition,

there is a group of 12 human genes with two mouse orthologs. For convenience, this group also is classified into category A. In category B, a single human gene is orthologous to a group of mouse genes (one-to-multiple orthology) and there are three such cases. The one-

to-one orthologous relationships are found in 13 gene pairs and these are classified into category C (one-to-one orthology). Five uncategorized human genes (*hT2R5*, *hT2R8*, *hT2R9*, *hT2R11\**, and *hT2R66\**) remain due to the absence of their counterparts in mice. Among 11 human pseudogenes, 6 (*hT2R15\**, *hT2R64\**, *hT2R63\**, *hT2R68\**, *hT2R65\**, and *hT2R67\**) occur in the 15 category A genes, none in the three category B genes, 3 (*hT2R2\**, *hT2R62\**, and *hT2R12\**) in the 13 category C genes, and 2 (*hT2R66\** and *hT2R11\**) in the 5 uncategorized genes. On the other hand, all 6 mouse pseudogenes (*mt2r24\**, *mt2r30\**, *mt2r32\**, *mt2r36\**, *mt2r37\**, and *mt2r40\**) are found in the 25 category B genes; none exist in the 3 category A genes and 13 category C genes.

**Primate *T2R* genes:** To elucidate the *T2R* gene repertoire in primates, we amplified 348 genes for three humans [Adygei (Eastern European), Japanese (Asian), Biaka (African Pygmies)], 12 nonhuman primates, and tupais by means of the PCR technique [33 genes from Hosa (Adygei); 32, Hosa (Japanese); 33, Hosa (Biaka); 36, Patr; 29, Gogo; 29, Popy; 26, Hyag; 35, Mamu; 26, Trcr; 20, Caja; 21, Ceap; 7, Otrc; 4, Gase; 5, Nyco; 8, Leca; and 4, Tugl]. The length of these amplified sequences ranges from 707 to 960 bp with the average being ~840 bp. As expected, the number of successfully amplified genes decreases in species distantly related to humans (such as prosimians and tupais) and this reduction is most likely due to mutations in the primer-attached sites. Among anthropoidea [hominoids, OWMs, and New World monkeys (NWMs)] in which at least 20 genes were obtained for each species, there are no significant differences in the proportion of *T2R* pseudogenes (Table 2, but see later results about differential rates of pseudogenization). However, the proportion differs greatly among the three categories of orthologous relationships. There are 46 pseudogenes among 115 genes in category A, none among 29 genes in category B, and 18 among 125 genes in category C. Fisher's exact test among pairs of categorized genes shows the preferential occurrence of pseudogenes in category A ( $P < 0.001$ ), or pseudogenes are statistically more abundant in category A genes with extra copies than in category B or C genes without extra copies. Nonetheless, it is worth keeping in mind that in primates, pseudogenization is not a rare event in genes without extra copies (category C genes).

To infer the evolutionary history of primate *T2R* genes, the NJ tree was again reconstructed on the basis of amino acid sequences of all available primate and mouse sequences (Figure 2). Since none of the primates have extra copies of category C genes, duplicability (the ability of loci to be duplicated) appears to be unchanged throughout the primate and mouse lineages. By contrast, duplicability in category A and B genes differs not only between primates and mice, but also within primates. In category B, the *T2R10* gene is one such

example. There are two *hT2R10* orthologs in slow loris (Nyco) and three in bush babies (Otrc) as well as five in mice, whereas there is only one *hT2R10* ortholog in nonhuman anthropoidea (Table 2). Duplicability at this locus in prosimians appears to be as high as in mice. In category A, on the other hand, the *T2R43* group (*T2R15*, -43-50, -63, -64, -68) experienced at least 19 duplications in anthropoidea (data not shown). This extensive duplication is in contrast to the case of mice in which only one gene duplication produced *mt2r17* and *mt2r23*, collectively orthologous to the primate *T2R43* group (Figures 1 and 2).

**Pseudogenization of *T2R* genes:** To determine a branch in the phylogenetic tree of primate *T2R* genes along which a particular gene was disrupted, we parsimoniously placed nonsense mutations as well as insertions and/or deletions (indels) that can cause frameshift mutations (Figure 2). Because there are no shared pseudogenes between humans and mice, no disruptive mutation was placed in the common ancestral lineage. Among primates, shared disruptive mutations result in eight pseudogenes common to more than one species (Figure 2). *T2R66* is an example of the pseudogenization that predated the divergence between anthropoidea and prosimians. The remaining 23 disruptive mutations occur on terminal branches and result in lineage-specific pseudogenes: 5 in prosimians and tupais, and 18 in anthropoidea (see below).

**Species-specific changes regarding pseudogenization:** Of 23 lineage-specific pseudogenes, 3 (*hT2R62\**, *hT2R2\**, and *hT2R64\**) are confined in humans (Figure 3). In *hT2R62\**, there are two nonsense mutations (CAG → TAG at codon position 235 and GAA → TAA at 292). Since both of these mutations are fixed in the sample from 17 human populations (Figure 3A) and since the chimpanzee ortholog is functional, *T2R62* in humans must have been inactivated long ago, but after the divergence between humans and chimpanzees. On the other hand, *hT2R2\** was produced relatively recently. It was previously annotated as a pseudogene owing to a two-base deletion at codon position 160 (CONTE *et al.* 2002; SHI *et al.* 2003). However, since we found the intact form in Adygei (Eastern European), Mbuti (African Pygmies), and Biaka (African Pygmies) (Figure 3B), the *hT2R2* locus is actually polymorphic in terms of the two-base deletion. Regarding *hT2R64\**, there is one fixed nonsense mutation of TGG → TGA at codon position 280. It is interesting to note that the orangutan ortholog (*Popy-T2R64\**) also suffers a nonsense mutation of TGG → TGA, but at codon position 49 (Figure 3C).

The remaining 19 lineage-specific pseudogenes are as follows: Patr (1), Gogo (2), Popy (1), Hyag (2), Mamu (2), Trcr (3), Caja (2), Ceap (1), Otrc (1), Leca (2), and Tugl (2). This observation indicates that all anthropoidea used in this study (Hosa, Patr, Gogo, Popy, Hyag, Mamu, Trcr, Caja, and Ceap) have their own pseudogenes. However, the rate of pseudogenization is high in the





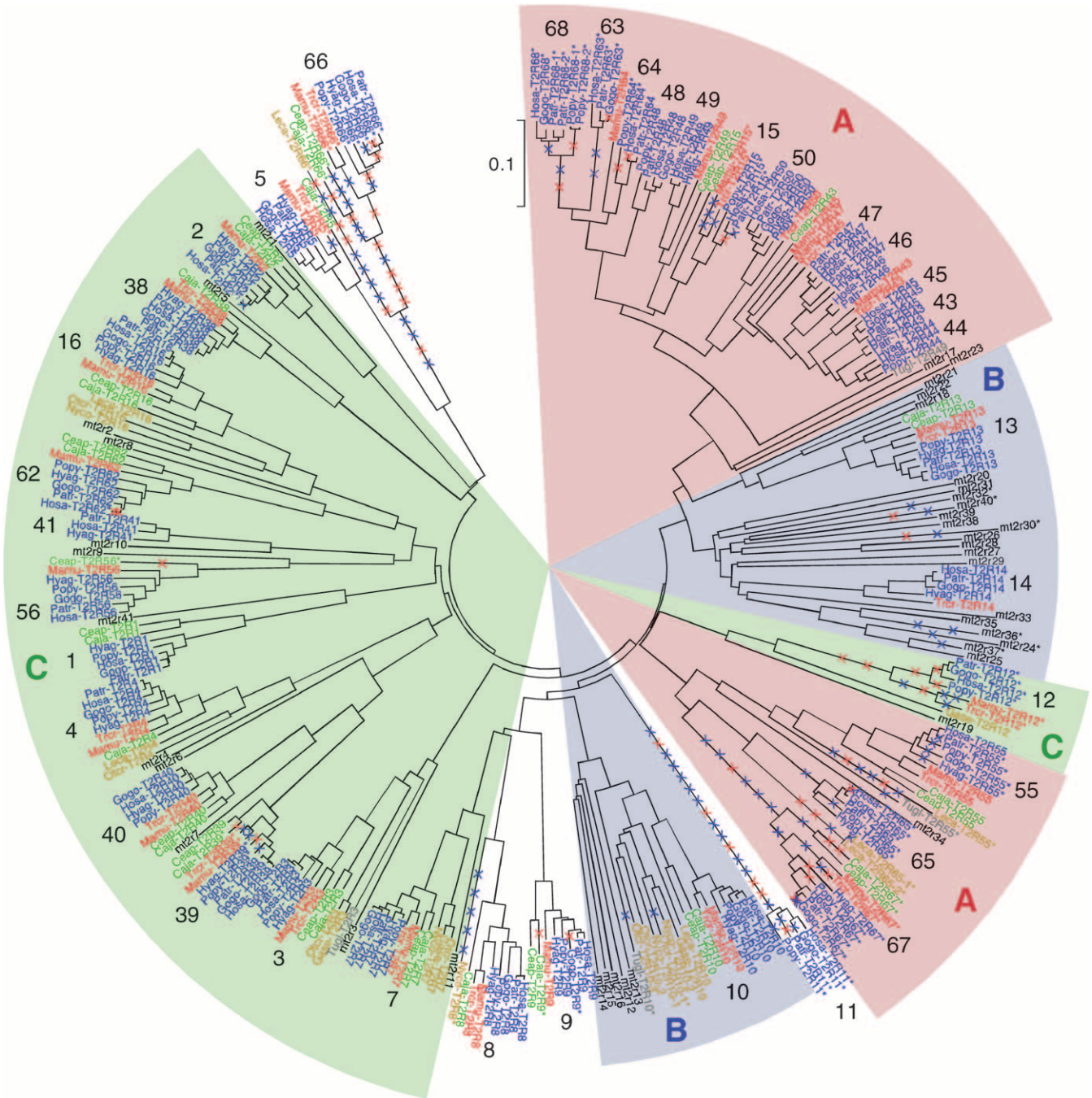


FIGURE 2.—Phylogenetic tree of 305 *T2R* genes based on amino acid sequences in primates and mice. Genes used are from hominoids (blue), OWMs (red), NWMs (green), prosimians (yellow), tupaia (gray), and mice (black). The NJ tree is reconstructed by using the  $p$  distances. The numbers in boldface type correspond to those in the locus designation system for human *T2R* genes. Pseudogenes are indicated by asterisks (\*). Nonsense (red crosses) and indel (blue crosses) mutations responsible for pseudogenization are placed along branches. The definitions of categories A, B, and C are the same as in Figure 1.

lineage leading to humans. Two fixed and one polymorphic pseudogene in humans arose within 6 million years (MY) after the divergence from chimpanzees, whereas 15 pseudogenes in nonhuman anthropoidea arose in a total of 117 MY (see GOODMAN *et al.* 1998 and Figure 4). The rate of pseudogenization is thus significantly different between humans and nonhuman anthropoidea (Fisher's exact test;  $P < 0.05$ , one-tailed). It is also interesting to note that disruptive mutations in a

particular gene occurred several times independently in different species. Like *T2R64* mentioned above, *T2R55* is disrupted independently in prosimians, tupaia, and four apes (chimpanzees, gorillas, orangutans, and agile gibbons), *T2R39* in two apes (orangutans and agile gibbons), two OWMs (rhesus monkeys and silvered leaf monkeys), and one NWM (marmosets), *T2R9* in gorillas and marmosets, and *T2R10* in bush babies and tree shrews (Figures 2 and 3D). In addition, *T2R15* be-



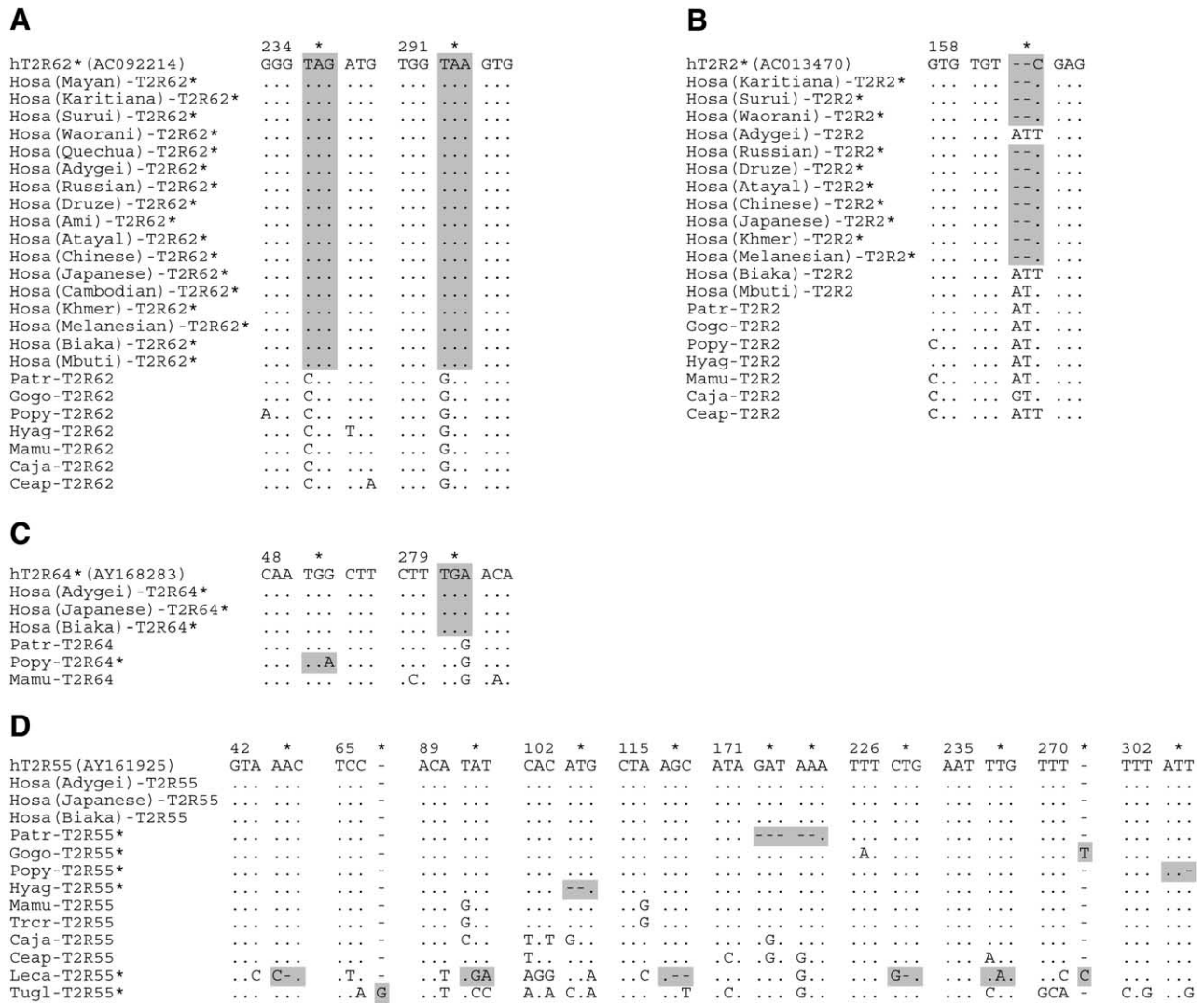


FIGURE 3.—Partial alignments of the nucleotide sequences of primate *T2R* genes. Dots show the same nucleotides as those in the top sequence, which were retrieved from GenBank. Both asterisks and shaded boxes show the location of mutations that impair intact ORFs. (A) *T2R62* is pseudogenized by two nonsense mutations (CAG → TAG at codon position 235 and GAA → TAA at 292). (B) *T2R2* is polymorphic in terms of a two-base deletion at codon position 160 in human populations. Only Adygei (from Eastern Europe), Biaka, and Mbuti Pygmies (both from Africa) possess the intact genes, while the deletion is found in all sequences collected from 10 other human populations (Karitiana, Surui, Waorani Indians from South America, Russians from Eastern Europe, Druze from the Middle East, Atayal, Chinese, Japanese from Eastern Asia, and Khmers and Melanesians from Southeast Asia) and from GenBank resources. (C) *T2R64* is a human-specific pseudogene (TGG → TGA at 280). This gene also became a pseudogene in the orangutan through an independent nonsense mutation (TGG → TGA at 49). (D) Independent mutations disrupt *T2R55* in the four apes: chimpanzees (five-base deletion at 172–173), gorillas (one-base insertion at 271), orangutans (one-base deletion at 303), and gibbons (two-base deletion at 103); in the prosimians: ring-tailed lemurs (two one-base deletions at 43 and 227, two-base deletion at 116, one-base insertion at 271, and two nonsense mutations at 90 and 236); and in the tupai: common tree shrews (one-base insertion at 66).

came nonfunctional by a single nonsense mutation in the common ancestral lineage between humans and chimpanzees as well as by independent indel mutations in rhesus monkeys and silvered leaf monkeys (Figure 2).

**Selection for and against functional domains of *T2R* genes:** We examined whether differential selective pressure has been acting on different domains of primate functional *T2R* genes. T2R molecules consist of three domains: transmembrane (TMs), intracellular (ICs), and extracellular (ECs) (ADLER *et al.* 2000). To evaluate selective pressure on each of these domains, we calcu-

lated the difference of  $d_N - d_S$  codon by codon (see MATERIALS AND METHODS) for 150 undisrupted sequences from primates and mice. To reduce sampling errors in this analysis, we excluded any codon that is absent in more than half the sequences (>75). Figure 5 shows that codons with  $d_N > d_S$  are abundant in ECs whereas codons with  $d_N < d_S$  predominate in TMs + ICs. Fisher's exact test ( $P < 0.01$ ) also shows that positively selected amino acids ( $d_N > d_S$ ) are much more frequently observed in ECs (15 of 84 sites) than in TMs + ICs (8 of 190) (see also SHI *et al.* 2003 for the comparison between

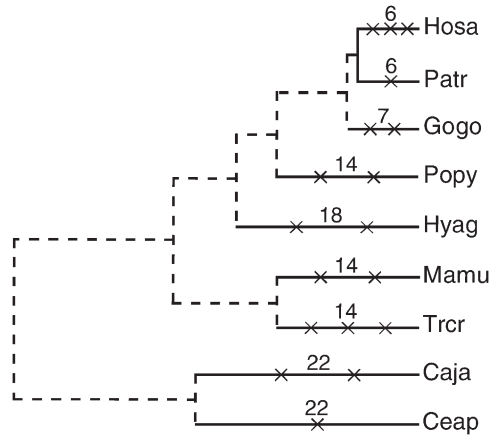


FIGURE 4.—The rate of pseudogenization in humans and nonhuman anthropoidea. Each cross on a terminal branch shows species-specific pseudogenization. The divergence times (MY) represented above each branch are from GOODMAN *et al.* (1998). While three human-specific pseudogenes have arisen within 6 MY, 15 (1 + 2 + 2 + 2 + 2 + 3 + 2 + 1) pseudogenes in nonhuman anthropoidea have accumulated within 117 MY (6 + 7 + 14 + 18 + 14 + 14 + 22 + 22). Pseudogenization that occurred along internodal branches (dashed lines) is excluded from consideration.

human and mouse *T2R* genes). Thus, some amino acid changes in ECs are selected for and most amino acid changes in TMs + ICs are selected against.

**Lineage-dependent selection on *T2R* genes:** It is also intriguing to examine how concordantly or discordantly selection operates between the primate and mouse lineages. We used 16 *T2R* genes in category B and C to analyze their  $b_s$  and  $b_N$  values (see MATERIALS AND METHODS) for 16 branches leading to primates (correspond-

ing to branch numbers from 1 to 16 in Table 3) and for 1 branch leading from the common node of primates to mice (corresponding to branch number 17 in Table 3). Among 272 possible comparisons of branches and genes in primates and mice, 227 pairs of the  $b_s$  and  $b_N$  values were obtained because of lack of sequences for some genes in some species. These comparisons were made separately for ECs (Table 3A) and TMs + ICs (Table 3B). For the primate EC comparisons, 8 show that  $b_N$  is significantly larger than  $b_s$  ( $b_N > b_s$ ), whereas for the 16 mouse comparisons, none show  $b_N > b_s$  but two exhibit  $b_s > b_N$ . The ratio of  $b_N$  to  $b_s$  in ECs averaged over all *T2R* genes is 0.94 in primates and 0.79 in mice. Although positive selection contributed to these relatively high ratios, relaxation of selective constraints in primates appears to be a more important factor. To test this conjecture, we compared the ratios in TMs + ICs between primates and mice. There is only 1 with  $b_N > b_s$  and 7 with  $b_s > b_N$  in the primate comparisons, whereas all 16 show  $b_s > b_N$  in the mouse comparisons. There is thus little evidence for positive selection in TMs + ICs and the average ratio of  $b_N$  to  $b_s$  is 0.57 in primates and 0.47 in mice. More importantly, since the increased amount of the ratio is similar in both domains, it is likely that primate *T2R* genes in categories B and C underwent relaxed negative selection rather than reinforced positive selection compared with the mouse orthologs.

Likewise, we examined the  $b_s$  and  $b_N$  values in primate category A genes for comparison. Figure 6 contrasts the  $b_s$  and  $b_N$  values between categories A and C as well as between domains EC and TM + IC. In both domains, the  $b_N$  value tends to be higher in category A than in category C irrespective of the  $b_s$  value. The Mann-Whitney

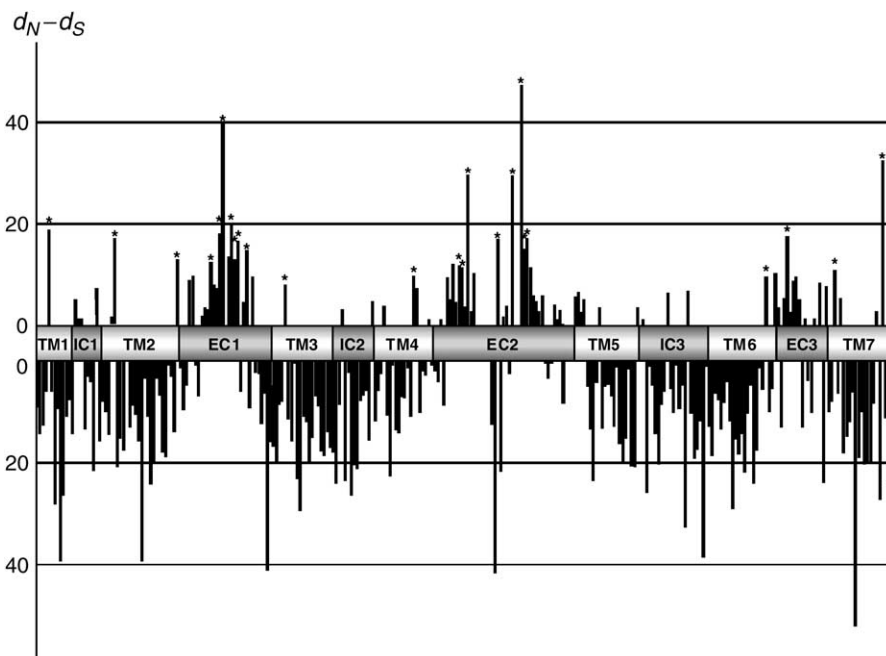


FIGURE 5.—The extent of positive and negative selection on individual codons. Schematic positions of three functional domains of the *T2R* gene (ECs, TMs, and ICs) are shown on the *x*-axis. The value above and below the *x*-axis, respectively, indicates an excess and deficiency in the number of per-site nonsynonymous substitutions ( $d_N$ ) over per-site synonymous ones ( $d_S$ ) in an individual codon. The zero value means  $d_N = d_S$ . Asterisks (\*) show significant positive selection ( $P < 0.05$ ) based on the extended binominal distribution (<http://www.datamonkey.org/>).

TABLE 3

Numbers of per-site synonymous substitutions ( $b_s$ ) and per-site nonsynonymous substitutions ( $b_n$ ) for each branch of the phylogenetic tree in category B and C genes

Out group	Hosa:	Patr:	Gogo:	Popy:	Hyag:	Mam:	Trcr:	Caja:	Ccap:	Mumu:							
no.:	1	2	3 <sup>a</sup>	4	5 <sup>a</sup>	6	7 <sup>a</sup>	8	9 <sup>a</sup>	10	11 <sup>a</sup>	12	13 <sup>a</sup>	14	15 <sup>a</sup>	16	17
A. Extracellular domains																	
<i>T2R1 mt2r41</i>	1.9/0.8	0.0/1.4	0.0/0.0	1.4/0.2	0.5/1.4	3.6/1.7	0.0/0.0	3.1/4.7					9.2/16.6	0.0/5.7#	1.1/3.2	0.4/4.9#	75.4/75.9
<i>T2R2 mt2r1</i>	1.9/0.0	0.0/1.5	0.0/0.0	0.0/2.3#	0.2/1.6	0.0/0.6	1.9/0.2	1.7/1.2	0.0/0.7	5.5/5.2			1.6/0.9	4.0/2.0	8.6/3.1	0.0/1.8	38.4/26.3
<i>T2R16 mt2r2</i>	0.0/0.6	0.0/0.0	2.2/0.1	0.0/1.0	0.0/0.6	1.9/1.6	0.1/0.6	0.0/1.1	1.4/0.6	0.0/3.1	0.0/4.9#	1.7/0.7	3.1/2.6	4.4/3.3	2.7/3.1	1.2/3.7	60.4/45.0
<i>T2R3 mt2r3</i>	1.5/0.6	0.0/0.0	0.0/0.0	0.0/0.0	0.3/0.3	2.8/0.7	0.0/0.0	1.4/2.2	0.0/2.6#	3.3/0.4	0.2/1.3	0.0/2.2#	0.0/3.1	9.1/2.4	8.8/6.2	5.2/1.0	32.0/20.6
<i>T2R4 mt2r4</i>	0.0/0.0	1.6/0.0	0.2/0.0	0.0/0.6	1.5/0.9	2.9/1.4	0.0/0.0	0.4/2.1	0.0/0.5	3.0/0.2	4.2/1.0	0.0/0.2	12.4/6.9	4.2/1.0			43.7/28.8
<i>T2R38 mt2r5</i>	1.5/0.0	1.4/0.0	0.0/0.0	0.3/0.0	0.4/0.3	1.9/1.6	0.0/0.0	3.2/2.4	0.0/1.0	0.0/0.4	3.6/0.8	1.4/1.1	2.9/4.4	6.8/4.3			35.7/24.4
<i>T2R39 mt2r6</i>	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0	0.2/0.6	1.4/0.0	0.0/0.0	3.4/2.9	0.0/1.1	3.3/0.2	2.3/2.3	0.0/0.8	0.0/2.8	3.0/0.6	6.2/2.8	0.2/1.0	46.8/55.4
<i>T2R40 mt2r7</i>	0.0/0.6			1.6/0.5	0.0/0.5	1.5/1.1	0.5/0.0	0.9/1.0	0.5/0.7	1.2/0.1	0.0/1.0	2.7/2.2	2.2/0.1	2.8/0.0	2.9/1.0	3.0/2.2	51.1/22.2*
<i>T2R62 mt2r8</i>	1.3/0.0	1.9/0.6	0.5/0.8	2.8/0.4	0.9/0.9	2.4/2.7	0.0/0.0	0.6/5.1#	3.2/1.5	0.0/0.0			0.0/0.0	4.5/7.0	14.6/6.4	16.0/10.8	62.7/62.4
<i>T2R56 mt2r9</i>	0.0/1.0	0.0/0.6	0.0/0.0	1.7/0.8	0.9/1.4	0.7/2.8	0.6/0.0	3.6/0.5	4.0/0.0	2.1/3.0			0.0/3.2		10.3/3.9		45.1/27.1
<i>T2R41 mt2r10</i>	0.0/0.2	0.0/1.4					0.6/1.3	3.7/2.8									60.0/28.4*
<i>T2R7 mt2r11</i>	0.0/0.0	1.4/0.0	0.0/0.2	0.0/0.4	2.2/0.0	0.1/1.5	0.0/0.0	4.8/0.5	3.1/2.5	2.5/0.9	0.0/1.4	1.4/0.0	0.9/0.0	5.2/1.7	2.9/2.7	4.3/0.8	22.4/19.9
<i>T2R10 mt2r12</i>	0.0/0.5	0.0/0.6	0.3/0.2	1.4/0.4	0.0/2.0	2.3/1.4	0.0/2.3	2.0/2.7	2.7/0.0	5.6/2.7	2.8/2.0	3.3/1.6	0.0/6.3	3.9/3.6	3.8/3.8	0.0/4.2#	59.3/54.8
<i>T2R12 mt2r19</i>	0.5/0.4	1.2/0.2	0.0/0.0	0.0/0.7	0.0/0.7	0.4/1.6			0.0/4.4	0.0/0.8	6.1/0.8	0.4/0.3					54.7/58.9
<i>T2R13 mt2r18</i>	0.0/0.5	1.9/1.7	0.0/0.0	0.0/0.0	0.0/0.7	0.0/1.5	0.0/0.2	0.0/1.0	2.3/2.7	1.8/1.9	3.4/2.6	0.0/0.9	5.7/1.8	3.0/3.8	1.4/9.8	1.8/4.9	41.3/45.0
<i>T2R14 mt2r27</i>	0.0/0.0	0.0/0.0	1.2/0.0	0.0/0.7	0.0/2.9	2.3/2.2	8.1/0.0				3.8/5.6						113.2/71.6
B. Transmembranes and intracellular domains																	
<i>T2R1 mt2r41</i>	0.0/0.3	0.6/0.0	0.2/0.0	1.8/0.0	4.2/0.7	2.1/0.7	0.8/0.2	2.1/2.1					2.1/1.1	4.4/1.6	8.6/3.6	2.7/1.5	45.2/23.1**
<i>T2R2 mt2r1</i>	0.0/0.0	0.1/0.2	0.0/0.0	0.0/1.0	0.0/0.3	3.2/0.2	0.3/0.0	0.0/0.7	1.2/1.1	4.3/1.8			2.3/0.0	5.8/1.3*	3.6/2.2	1.1/1.2	44.9/20.0**
<i>T2R16 mt2r2</i>	0.6/0.3	0.0/0.0	1.2/0.0	1.5/0.0	0.5/0.3	0.7/1.3	0.6/0.0	1.4/1.1	1.1/1.2	1.9/1.0	1.3/0.6		0.5/1.6	1.5/3.5	2.5/2.7	3.2/2.3	46.2/22.6**
<i>T2R3 mt2r3</i>	0.0/0.0	0.6/0.0	0.0/0.0	0.0/0.0	1.9/0.8	0.9/1.1	0.5/0.0	1.2/0.3	0.0/0.7	3.4/0.9	2.5/1.2	1.1/1.3	2.3/2.1	2.1/0.2	2.6/2.6	3.8/1.7	33.3/18.5*
<i>T2R4 mt2r4</i>	0.0/0.2	0.0/0.3	0.1/0.0	0.5/0.7	1.1/0.2	2.2/0.6	0.0/0.0	1.9/1.1	1.4/0.5	0.9/0.8	1.0/0.8	0.5/0.5	0.2/1.5	3.4/2.7			33.5/13.8***
<i>T2R38 mt2r5</i>	0.0/0.6	0.6/0.1	0.0/0.0	0.0/0.3	1.0/0.2	0.7/0.6	0.1/0.1	2.8/1.0	1.6/0.0	0.4/0.3	1.7/0.2	3.0/0.7	2.9/1.4	3.3/3.1			38.6/16.1***
<i>T2R39 mt2r6</i>	0.0/0.0	0.5/0.5	0.0/0.2	1.2/0.0	0.9/0.3	0.9/0.7	0.0/0.6	0.9/0.6	1.3/0.1	1.4/0.3	1.7/1.4	2.9/0.6	2.3/0.3	3.0/1.0	1.1/2.0	3.8/0.2*	38.5/19.2**
<i>T2R40 mt2r7</i>	0.0/0.0			0.0/0.0	0.0/0.0	1.4/0.2	0.0/0.0	0.0/0.2	0.7/0.5	0.9/0.3	0.0/0.6	2.5/1.4	1.4/0.3	0.8/1.6	1.5/1.3	4.7/1.1	27.0/14.7*
<i>T2R62 mt2r8</i>	1.1/0.6	0.7/0.3	0.0/0.0	0.9/0.2	1.2/0.6	2.8/1.4	0.4/0.3	0.9/1.1	3.7/2.4	0.0/0.0			0.0/0.0	4.6/3.4	10.7/3.8*	12.4/5.1*	43.2/25.5*
<i>T2R56 mt2r9</i>	0.0/0.5	1.2/0.7	0.7/0.3	1.0/0.7	1.0/1.7	3.1/1.9	0.0/0.0	3.0/1.3	1.9/0.0	2.2/2.5			5.5/1.6		10.7/6.9		43.7/22.8**
<i>T2R41 mt2r10</i>	0.5/0.5	0.6/0.0					1.5/0.3	0.8/1.6									43.5/17.0***
<i>T2R7 mt2r11</i>	0.0/0.2	0.7/0.6	0.0/0.0	0.8/0.3	0.9/0.3	2.9/0.4	0.0/0.0	2.2/1.1	0.9/0.3	3.1/1.0	1.9/0.4	4.4/1.0	0.6/1.0	3.3/1.7	2.1/1.6	5.1/1.0*	28.6/15.3*
<i>T2R10 mt2r12</i>	0.0/0.5	0.0/0.0	0.0/0.0	0.0/0.6	1.5/0.5	2.5/0.8	0.4/0.0	2.1/1.5	0.0/1.5	0.7/0.5	6.1/0.6*	0.8/1.2	0.0/2.9	5.9/3.3	9.9/5.3	2.0/1.3	51.6/18.6***
<i>T2R12 mt2r19</i>	1.7/0.0	1.9/0.3	0.0/0.3	0.6/0.0	0.3/0.3	1.6/0.9			4.6/0.3	0.9/2.7	0.0/6.1#	0.9/2.5					62.2/31.2**
<i>T2R13 mt2r18</i>	0.0/0.9	1.2/0.9	0.0/0.0	0.0/0.9	1.8/0.9	1.1/2.1	0.0/0.2	1.0/0.6	2.5/1.0	1.1/1.6	1.2/1.0	0.7/0.5	1.3/0.3	1.4/2.0	0.7/0.7	11.2/1.6**	52.8/16.1***
<i>T2R14 mt2r27</i>	1.4/0.4	0.1/0.7	1.4/0.6	1.2/0.0	0.5/2.1	1.3/1.9	2.8/1.5				4.6/4.4						58.4/29.0**

The  $b_N \times 100$  and  $b_N \times 100$  values are given before and after the “/” in (A) extracellular domains (ECs) and (B) transmembranes and intracellular domains (TMs + ICs).  $b_S > b_N$  at significant levels of 0.05 (\*), 0.01 (\*\*), 0.001 (\*\*\*) and  $b_N > b_S$  at a significant level of 0.05 (#) in the z-test. Pseudogenes are underlined.

<sup>a</sup>The branch number 3 means the branch from the node of ((Hosa, Patr), Gogo) to the node of ((Hosa, Patr), Gogo), Popy) to the node of ((Hosa, Patr), Gogo), and so on.

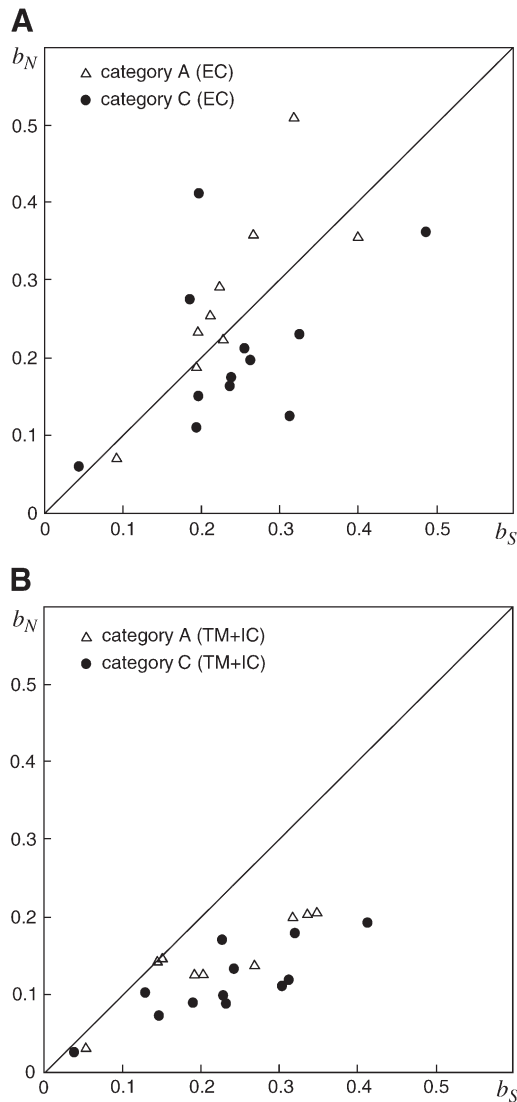


FIGURE 6.—The  $b_S$  and  $b_N$  values of category A genes ( $\Delta$ ) and category C genes ( $\bullet$ ) in ECs (top) and TMs + ICs (bottom). The  $x$ - and  $y$ -axis stand for  $b_S$  and  $b_N$ , respectively. The primate genes used for category A are *T2R15*, *T2R43-44*, *T2R45*, *T2R46*, *T2R47*, *T2R48*, *T2R49*, *T2R50*, and *T2R55*, and those for category C are *T2R1*, *T2R2*, *T2R3*, *T2R4*, *T2R7*, *T2R16*, *T2R38*, *T2R39*, *T2R40*, *T2R41*, *T2R56*, and *T2R62*.

test also revealed that the ratio of  $b_N$  to  $b_S$  is significantly higher in category A genes than in category C genes in both ECs ( $P < 0.05$ ) and TMs + ICs ( $P < 0.01$ ). Similar to the comparison between primates and mice, these results indicate that category A genes within primates underwent relaxed negative selection.

## DISCUSSION

In this article, we have characterized the repertoire of bitter taste receptor (*T2R*) genes in humans and non-human primates to understand the tempo and mode of the duplication and pseudogenization process. It is

generally difficult to discern the complete repertoire of other chemosensory receptor genes in mammals. In the case of olfactory receptor (*OR*) genes, this difficulty is due mainly to the fact that they form an exceptionally large multigene family (856 genes in humans, 1404 in mice, and 971 in dogs according to the HORDE database at <http://bip.weizmann.ac.il/HORDE/>). On the other hand, the number of *T2R* genes is moderately small and suspected to be at most  $\sim 40$  in the mammalian genome. This fact has facilitated our study to discern the great majority of the *T2R* gene repertoire in anthropoidea. Indeed, for each nonhuman anthropoid we have identified at least 20 genes that are orthologous to 36 human *T2R* genes (Table 2).

In *OR* genes, hominoids (especially humans) and OWMs have a higher proportion of pseudogenes than do NWMs or lemurs (GILAD *et al.* 2003, 2004), and in pheromone receptor (*VR*) genes, hominoids and OWMs are thought to have completely lost the functional genes (ZHANG and WEBB 2003). By contrast, among anthropoidea there are no significant differences in the proportion of the *T2R* pseudogenes, which suggests that the tempo and mode of pseudogenization in *T2R* is different from that in *OR* and *VR*.

In primates, many more pseudogenes have accumulated in category A (46 pseudogenes of 115 genes) than in category B (0 of 29) and in category C (18 of 125) (Fisher's exact test;  $P < 0.001$ ). In mice, all pseudogenes are restricted to category B and absent in categories A and C. This contrast strongly suggests that the more frequently duplication occurs, the more pseudogenes tend to accumulate. This trend is also evident for prosimians, in which three duplicated *T2R10* genes are found and one of them is a pseudogene (Figure 2 and Table 2). It therefore appears that the evolutionary process of *T2R* genes exemplifies the birth-and-death model for multigene families (NEI *et al.* 1997), as for *MHC* (HUGHES and NEI 1989; NEI *et al.* 1997), immunoglobulin (OTA and NEI 1994), histone (PIONTKIVSKA *et al.* 2002), and *OR* genes (YOUNG *et al.* 2002). However, major determinants that have shaped the gene repertoire may well differ from family to family. Broadly, determinants are either internal (genomic) or external (environmental). The gene repertoire of rRNA and histone multigene families is likely to be determined by internal (physiological or genomic) requirements. On the other hand, the gene repertoire in *MHC* and *OR* genes is likely to be determined by requirements for defense against exogenous pathogens and perception of chemical substances, respectively. The situation of *T2R* resembles that of *MHC* and *OR* in that the gene repertoire is largely shaped by available foods.

Because *T2R* genes in category C are conserved to maintain the one-to-one orthologous relationships between primates and mice, we can speculate that *T2R* in category C can perceive bitter substances common to mammals. In this regard, it is worth noting that all three

human-specific pseudogenes (*hT2R2\**, *hT2R12\**, and *hT2R62\**) belong to category C (Figure 2 and Table 1). Equally importantly, humans have accumulated more pseudogenes per unit of time than nonhuman anthropoidea (Figure 4). Taken together, these observations indicate accelerated deterioration of tasting particular bitter substances in humans. Interestingly, there are also lineage-specific pseudogenes in some nonhuman primates and such pseudogenization tends to repeat in different lineages. As shown in Figure 2, there are six such genes (*T2R64*, *T2R55*, *T2R39*, *T2R9*, *T2R10*, and *T2R15*) that were repeatedly dysfunctional in more than one primate species. Repeated disruptions of these genes in different primates and tupais strongly suggest that the fate of taste receptor genes is associated with environmental factors rather than with genetic ones. Nevertheless, since *T2R55* and *T2R39* are functional in humans, these genes are likely indispensable to tasting particular bitter substances throughout human evolution.

We have demonstrated that functional constraints are more relaxed for the primate *T2R* genes than for the mouse genes (Table 3). Since the number of positively selected amino acids in ECs is significantly larger than that in TMs + ICs (Figure 5), the average ratio of  $b_N$  to  $b_S$  in primates is greater in ECs (0.94) than in TMs + ICs (0.57). Similarly, the average ratio in mice is 0.79 in ECs and 0.47 in TMs + ICs. We have noted that increased amounts of these ratios in primates relative to mice are almost the same between ECs (0.94/0.79) and TMs + ICs (0.57/0.47), indicating that irrespective of the domains, ~20% of the functional constraints are more relaxed in the primate lineage than in the mouse lineage. This relaxation in the primate lineage is reflected in the presence of a relatively larger number of pseudogenes in primates (20~35%) than in mice (15%). Recently, two research groups studied the pattern of the molecular evolution of *T2R* genes in humans and chimpanzees (PARRY *et al.* 2004; WANG *et al.* 2004). Specifically, WANG *et al.* (2004) examined the extent of functional constraints in 25 pairs of human and chimpanzee *T2R* orthologs and in 28 pairs of mouse and rat orthologs. Their conclusion is similar to ours: The *T2R* genes in humans and chimpanzees are under relaxed selective constraints compared with those in mice and rats. Furthermore, FISCHER *et al.* (2005) reported the evolutionary study of *T2R* genes in humans, four great apes, and two OWMs and compared the ratio of per-site nonsynonymous substitutions to per-site synonymous substitutions in primates ( $f_p$ ) with that in rodents ( $f_r$ ). Since the  $f_p$  value is significantly larger than  $f_r$ , these authors argued that the elevated ratio in primates results from the reduced effective population size and can be attributed to fixation of slightly deleterious mutations (OHTA 1975). Theoretically speaking, if the reduced effective population size is truly the cause of the elevated ratio in primates, all *T2R* genes should be equally influenced by increased amounts of effectively

neutral mutations in a small population. This conclusion can be tested by the  $f$  values of individual *T2R* genes between primates and mice. Under the slightly deleterious hypothesis, we would expect more or less the same ratio of  $f_p$  to  $f_r$  regardless of the genes involved. However, this is not the case. Although the  $f_p/f_r$  ratio averaged over 31 genes is 1.27 and consistent with the relatively small effective size in primates, the individual ratio values vary greatly from 0.58 (*T2R15*) to 2.21 (*T2R41*). This large variation in the  $f_p/f_r$  ratio among loci can be more easily accounted for by relaxed negative selection than by reduced effective size.

We have shown that functional constraints are weaker against genes in category A than in category C (Figure 6). Hence, there is a trend that newly duplicated category A genes in primates tend to have accumulated non-synonymous substitutions and contributed to expansion of the gene repertoire, which might allow primates to perceive numerous bitter substances in a changing environment. We have also shown that primate *T2R* genes were subjected to extensive lineage-dependent pseudogenization of functionally diversified duplicated copies. It appears that under joint effects of gene duplication and pseudogenization as well as interplay between genes and environments, primate *T2R* genes have evolved toward species-specific repertoires.

This research was supported in part by the Japanese Society for Promotion of Science grant no. 12304046 (to N.T.).

#### LITERATURE CITED

- ADLER, E., M. A. HOON, K. L. MUELLER, J. CHANDRASHEKAR, N. J. RYBA *et al.*, 2000 A novel family of mammalian taste receptors. *Cell* **100**: 693–702.
- CHANDRASHEKAR, J., K. L. MUELLER, M. A. HOON, E. ADLER, L. FENG *et al.*, 2000 T2Rs function as bitter taste receptors. *Cell* **100**: 703–711.
- CONTE, C., M. EBELING, A. MARCUZ, P. NEF and P. J. ANDRES-BARQUIN, 2002 Identification and characterization of human taste receptor genes belonging to the TAS2R family. *Cytogenet. Genome Res.* **98**: 45–53.
- CONTE, C., M. EBELING, A. MARCUZ, P. NEF and P. J. ANDRES-BARQUIN, 2003 Evolutionary relationships of the Tas2r receptor gene families in mouse and human. *Physiol. Genomics* **14**: 73–82.
- FISCHER, A., Y. GILAD, O. MAN and S. PÄÄBO, 2005 Evolution of bitter taste receptors in humans and apes. *Mol. Biol. Evol.* **22**: 432–436.
- FITCH, W. M., 1971 Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool.* **20**: 406–416.
- GILAD, Y., O. MAN, S. PÄÄBO and D. LANCET, 2003 Human specific loss of olfactory receptor genes. *Proc. Natl. Acad. Sci. USA* **100**: 3324–3327.
- GILAD, Y., V. WIEBE, M. PRZEWORSKI, D. LANCET and S. PÄÄBO, 2004 Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol.* **2**: 120–125.
- GOODMAN, M., C. A. PORTER, J. CZELUSNIAK, S. L. PAGE, H. SCHNEIDER *et al.*, 1998 Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol. Phylogenet. Evol.* **9**: 585–598.
- HARTIGAN, J. A., 1973 Minimum evolution fits to a given tree. *Biometrics* **29**: 53–65.
- HILLIER, L. W., W. MILLER, E. BIRNEY, W. WARREN, R. C. HARDISON *et al.*, 2004 Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**: 695–716.

- HUGHES, A. L., and M. NEI, 1989 Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. *Mol. Biol. Evol.* **6**: 559–579.
- KIM, U. K., E. JORGENSEN, H. COON, M. LEPPERT, N. RISCH *et al.*, 2003 Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science* **299**: 1221–1225.
- KOSAKOVSKY POND, S. L., and S. D. W. FROST, 2005 Not so different after all: comparison of various methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **22**: 1208–1222.
- KUMAR, S., K. TAMURA, I. B. JAKOBSEN and M. NEI, 2001 MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**: 1244–1245.
- LEHNINGER, A. L., 1996 *Biochemistry*, Ed. 2. Kalyani Publishers, New Delhi.
- LI, X., L. STASZEWSKI, H. XU, K. DURICK, M. ZOLLER *et al.*, 2002 Human receptors for sweet and umami taste. *Proc. Natl. Acad. Sci. USA* **99**: 4692–4696.
- MATSUNAMI, H., and H. AMREIN, 2003 Taste and pheromone perception in mammals and flies. *Genome Biol.* **4**: 220.
- MATSUNAMI, H., J. P. MONTMAYEUR and L. B. BUCK, 2000 A family of candidate taste receptors in human and mouse. *Nature* **404**: 601–614.
- MUSE, S. V., and B. S. GAUT, 1994 A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol. Biol. Evol.* **11**: 715–724.
- NEI, M., X. GU and T. SITNIKOVA, 1997 Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. USA* **94**: 7799–7806.
- NELSON, G., M. A. HOON, J. CHANDRASHEKAR, Y. ZHANG, N. J. RYBA *et al.*, 2001 Mammalian sweet taste receptors. *Cell* **106**: 381–390.
- NISHIKIMI, M., R. FUKUYAMA, S. MINOSHIMA, N. SHIMIZU and K. YAGI, 1994 Cloning and chromosomal mapping of the human nonfunctional gene for L-gulonolactone oxidase, the enzyme for L-ascorbic acid biosynthesis missing in man. *J. Biol. Chem.* **269**: 13685–13688.
- OHNO, S., 1970 *Evolution by Gene Duplication*. Springer-Verlag, New York.
- OHTA, T., 1975 Statistical analyses of Drosophila and human protein polymorphisms. *Proc. Natl. Acad. Sci. USA* **72**: 3194–3196.
- OTA, T., and M. NEI, 1994 Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Mol. Biol. Evol.* **11**: 469–482.
- PARRY, C. M., A. ERKNER and J. LE COUTRE, 2004 Divergence of T2R chemosensory receptor families in humans, bonobos, and chimpanzees. *Proc. Natl. Acad. Sci. USA* **101**: 14830–14834.
- PIONTKIVSKA, H., A. P. ROONEY and M. NEI, 2002 Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol. Biol. Evol.* **19**: 689–697.
- SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- SHI, P., J. ZHANG, H. YANG and Y. P. ZHANG, 2003 Adaptive diversification of bitter taste receptor genes in mammalian evolution. *Mol. Biol. Evol.* **20**: 805–814.
- SUZUKI, Y., and T. GOJOBORI, 1999 A method for detecting positive selection at single amino acid sites. *Mol. Biol. Evol.* **16**: 1315–1328.
- THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN and D. G. HIGGINS, 1997 The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**: 4876–4882.
- WANG, X., S. D. THOMAS and J. ZHANG, 2004 Relaxation of selective constraint and loss of function in the evolution of human bitter taste receptor genes. *Hum. Mol. Genet.* **13**: 2671–2678.
- WONG, G. T., K. S. GANNON and R. F. MARGOLSKEE, 1996 Transduction of bitter and sweet taste by gustducin. *Nature* **381**: 796–800.
- YOUNG, J. M., C. FRIEDMAN, E. M. WILLIAMS, J. A. ROSS, L. TONNES-PRIDDY *et al.*, 2002 Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Hum. Mol. Genet.* **11**: 535–546.
- ZHANG, J., and D. M. WEBB, 2003 Evolutionary deterioration of the vomeronasal pheromone transduction pathway in catarrhine primates. *Proc. Natl. Acad. Sci. USA* **100**: 8337–8341.
- ZHANG, J., H. ROSENBERG and M. NEI, 1998 Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA* **95**: 3708–3713.

Communicating editor: S. SHAEFFER