J Mol Evol (1993) 36:249–254

Journal of Molecular Evolution

Adaptive Evolution in the Rat Olfactory Receptor Gene Family

Austin L. Hughes and Marianne K. Hughes

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

Summary. Comparison of DNA sequences of the rat (*Rattus norvegicus*) olfactory receptor gene family revealed an unusual pattern of nucleotide substitution in the gene region encoding the second extracellular domain (E2) of the protein. In this domain, nonsynonymous nucleotide differences between members of this subfamily that caused a change in amino acid residue polarity were over four times more frequent than nonsynonymous differences that did not cause a polarity change. This nonrandom pattern of nucleotide substitution is evidence of past directional selection favoring diversification of the E2 domain among members of this subfamily. This in turn suggests that E2 may play some important role in the functions unique to each member of the olfactory receptor family, and that it may perhaps be an odorant binding domain.

Key words: Adaptive evolution — Olfactory receptor — Multigene families

The duplication of genes, leading to the formation of families of genes whose members have distinct but related functions, is fundamental to adaptive evolution (Ohno et al. 1968; Li 1983). However, it is not well understood how such duplicated genes evolve new functions. One hypothesis is that after gene duplication one gene copy is redundant and thus is free to accumulate nonsynonymous nucleotide substitutions in positions where such substitutions would be deleterious to the functional protein. These replacements then fortuitously adapt the product of this gene to a new function (Kimura and Ohta 1974). Alternatively, it has been argued that gene duplication may be followed by a burst of amino acid replacements fixed by directional selection that adapt the duplicated locus to a different function from that of the parent locus (Goodman et al. 1975). In order to decide which of these two models of evolution is applicable to a family of rat (*Rattus norvegicus*) olfactory receptor proteins (Buck and Axel 1991), we examined rates of nucleotide substitution in different regions of genes belonging to this family.

Functionally important regions of proteins are usually identified by conservation of amino acid sequence. However, when members of a multigene family have diverged functionally, regions important to the specific function of each protein will not be conserved across all members of the family. Thus, it is difficult to identify such regions by sequence comparison alone. If there is evidence that a region has undergone past adaptive evolution leading to diversification among members of a gene family, such evidence suggests that the region involved may be important to the specific function of each family member. Here we present such evidence in the case of one of the extracellular domains of the rat olfactory receptor. The evidence of past adaptive evolution in this region in turn identifies it as a candidate odorant binding domain.

DNA Sequences Analyzed

The rat olfactory receptors are members of the G protein–coupled receptor or heptahelical superfamily (O'Dowd et al. 1989; Mollon 1991). Members of

Offprint requests to: A.L. Hughes

this family are transmembrane proteins with the N-terminus extracellular and the C-terminus intracellular; there are four extracellular domains (E1–E4), seven transmembrane domains (T1–T7), and four intracellular domains (I1–I4). For the 10 cDNA sequences available for members of this family from the rat, we computed proportions of amino acid differences and numbers of nucleotide substitutions per site in pairwise comparisons. The sequences were aligned following Buck and Axel (1991); so that a comparable data set was used in each pairwise comparison, codons where the alignment postulated a gap were excluded from all comparisons. The numbers of codons analyzed in each domain were as follows: E1 (23); E2 (21); E3 (38); E4 (19); T1 (25); T2 (22); T3 (20); T4 (19); T5 (23); T6 (24); T7 (21); I1 (6); I2 (18); I3 (17); I4 (16).

Results

Phylogenetic Analysis

Phylognentic trees of rat olfactory receptor genes were constructed by the neighbor-joining method (Saitou and Nei 1987) based on the number of nonsynonymous substitutions per site (d_N), which was estimated by Nei and Gojobori's (1986) method. In a preliminary analysis, separate phylogenetic trees were constructed separately for the 15 different domains in order to determine whether different gene regions have different phylogenetic histories. The only domain which differed markedly from others was I3. Here, we present two phylogenetic trees: a tree based on all domains except I3 (Fig. 1A) and a tree based on I3 (Fig. 1B). In domains other than I3, the rat olfactory proteins could be separated into three subfamilies, which we designate I, II, and III. The lengths of the internal branches separating these three subfamilies are all significantly nonzero (Fig. 1A). Subfamily III has the most representatives (six) among the available sequences; and, as indicated by the fact that branch lengths within the subfamily III cluster are shorter than branch lengths within the other two subfamilies (Fig. 1A), the available sequences from subfamily III are more closely related to each other than are members of the other two subfamilies.

In I3, the subfamily III gene F5 clusters with subfamily I. This clustering is not statistically significant, but the number of nucleotide sites in this domain is quite small. If it is not simply the result of chance, the resemblance of F5 to subfamily I in I3 might be explained by either of these two hypotheses: (1) an interlocus recombinational event, involving a mechanism such as exon-shuffling or gene conversion, has given F5 an I3 sequence originating from subfamily I; or (2) F5 has come to resemble subfamily I in the I3 region through convergent evolution. On the former hypothesis but not the latter, F5 would be expected to resemble subfamily I sequences at synonymous sites as well as nonsynonymous sites. Unfortunately, synonymous sites are almost saturated in comparisons among these genes; thus the number of synonymous substitutions per site cannot be estimated accurately for most comparisons among these genes (data not shown). Thus it is not at present possible to decide between these two hypotheses.

Adaptive Evolution of Protein Domains

Table 1 shows mean percent amino acid difference (p) in the 15 domains for comparisons within and among the three subfamilies. None of the four extracellular domains consistently showed any difference in extent of conservation from the other regions analyzed (Table 1). Likewise, none of the intracellular domains was consistently more conserved than the others, while among the transmembrane domains, the most consistent tendency was to conserve T2 (Table 1). The consistent conservation of T2 both within and between subfamilies suggests that this domain may play an important functional role common to all members of the family.

In order to test further for differences among regions, we applied a method of examining the extent to which amino acid properties are conserved (Hughes et al. 1990). This method divides nonsynonymous (amino-acid-altering) nucleotide differences into those that are conservative and those that are radical (nonconservative) with respect to an amino acid residue property of interest. Each nonsynonymous site (or fractional site, as defined by Nei and Gojobori 1986) is classified as a conservative or radical site (or some fraction conservative and some fraction radical), and the number of conservative nonsynonymous differences per site (p_{NC}) and the number of radical nonsynonymous differences per site (p_{NR}) are calculated. If in a given region $p_{NC} > p_{NR}$, amino acid replacements have occurred in such a way as to conserve the amino acid property. On the other hand, if $p_{NR} > p_{NC}$, amino acid replacements causing a change in the amino acid property have occurred at a greater rate than expected under random substitution; thus, $p_{NR} > p_{NC}$ is evidence of positive selection acting to change the amino acid properties of a protein region.

The presence of polar groups characterizes many odorants known from behavioral data to be recognized by mammals (Wheeler 1977, Beets 1978). Thus, we reasoned that interaction of olfactory re-

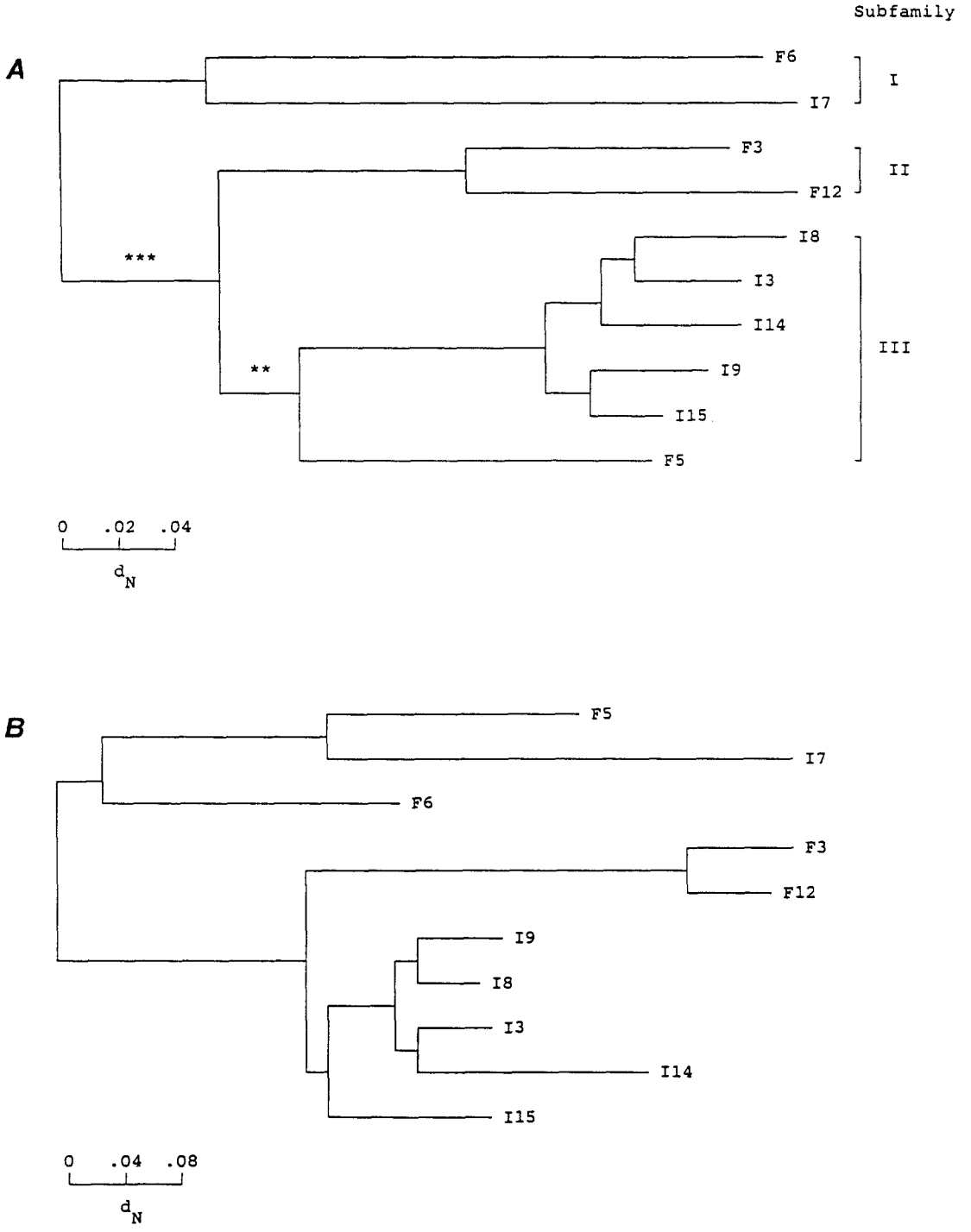

Fig. 1. **A** Phylogenetic tree of coding regions (excluding the I3 domain) of 10 cDNAs sequences from rat olfactory receptor genes based on number of nonsynonymous nucleotide substitutions per site site (d_N) in 287 aligned codons. Tests of significance of internal branches (Li 1989): $^*P < .05$; $^{**}P < .01$; $^{***}P < .001$. **B** Phylogenetic tree of I3 domain from the same genes (17 aligned codons).

ceptors with different volatile compounds might involve patterns of residue polarity. We computed mean p_{NC} and p_{NR} with respect to residue polarity in comparisons among and within subfamilies of rat olfactory protein genes (Table 2). In E2, in the comparison among subfamily III sequences, p_{NR} was over four times p_{NC}, and the difference between the two was highly significant statistically (Table 2). In other comparisons, p_{NR} in E2 was not significantly different from p_{NC}. However, p_{NR} in E2 was always considerably higher than p_{NR} in the other extracellular domains or in the transmembrane or intracellular domains, whereas p_{NC} in E2 was generally very similar to p_{NC} in these other regions (Table 2). On the other hand, in E1, E3, and E4 and in the transmembrane and intracellular portions, p_{NC} always exceeded p_{NR}, and this difference was statistically significant in a majority of cases (Table 2).

Table 1. Mean percent amino acid difference (p) (±S.E.) in comparisons of different domains in three subfamilies of rat olfactory proteins[a]

Domain	I vs I	I vs II	I vs III	II vs II	II vs III	III vs III
E1	65.2 ± 9.9*	67.4 ± 8.5*	70.7 ± 9.3	56.5 ± 10.3***	68.5 ± 9.5	38.8 ± 6.7***
E2	61.9 ± 10.6*	66.7 ± 10.1*	65.5 ± 9.7	33.3 ± 10.3	54.4 ± 9.8***	27.4 ± 5.1**
E3	42.1 ± 8.0	71.1 ± 7.7**	49.8 ± 7.0	42.1 ± 8.0**	63.4 ± 7.7*	27.4 ± 5.1**
E4	76.9 ± 11.7**	86.6 ± 12.2**	85.3 ± 12.0*	53.9 ± 13.8**	74.4 ± 12.5**	41.5 ± 8.2***
T1	64.0 ± 9.6*	56.0 ± 8.9	55.0 ± 7.9	28.0 ± 9.0	43.0 ± 7.8**	30.9 ± 6.2**
T2	31.8 ± 9.9	38.6 ± 9.3	50.0 ± 10.1	13.6 ± 7.3	27.7 ± 7.8	10.3 ± 3.9
T3	30.0 ± 10.2	56.3 ± 10.5	50.8 ± 9.7	40.0 ± 11.0*	45.0 ± 7.8	36.3 ± 7.5**
T4	57.9 ± 11.3	76.3 ± 10.6**	76.3 ± 10.5	52.6 ± 11.5**	56.6 ± 8.9*	48.1 ± 8.3***
T5	65.2 ± 9.9*	80.4 ± 9.4**	75.4 ± 9.4	56.5 ± 10.3***	71.0 ± 9.5***	44.9 ± 7.1***
T6	37.5 ± 9.9	49.0 ± 9.6	46.2 ± 0.9	12.5 ± 6.8	30.9 ± 8.0	16.1 ± 4.7
T7	57.1 ± 10.8	48.8 ± 9.9	52.8 ± 9.4	9.5 ± 6.4	25.3 ± 7.7	16.2 ± 5.2
I1	66.7 ± 19.2	58.3 ± 17.2	68.1 ± 18.9	50.0 ± 20.4	25.4 ± 7.7	16.2 ± 5.2
I2	50.0 ± 11.8	38.9 ± 8.5	38.9 ± 8.9	27.8 ± 10.6	36.6 ± 9.2	22.6 ± 6.2
I3	52.9 ± 12.1	73.4 ± 11.6*	66.7 ± 10.1	23.5 ± 10.3	62.7 ± 11.3*	62.7 ± 11.3***
I4	37.5 ± 12.1	43.8 ± 9.5	52.6 ± 10.8	50.0 ± 12.5*	56.3 ± 10.6*	56.3 ± 10.6***

[a] Standard errors were computed by Nei and Jin's (1989) method. Tests of significance of the difference between p and that in T2: $*P < .05$; $**P < .01$; $***P < .001$

Table 2. Mean percent conservative (p_{NC}) and radical (p_{NR}) nonsynonymous nucleotide difference with respect to amino acid residue polarity (±S.E.), in comparisons among rat olfactory protein gene subfamilies[a]

	E2		E1, E3, E4	
Comparison	p_{NC}	p_{NR}	p_{NC}	p_{NR}
I vs I	39.8 ± 9.0	45.1 ± 11.8	39.1 ± 6.4	23.8 ± 7.9
vs II	46.2 ± 7.8	34.1 ± 10.1	36.5 ± 4.6	17.1 ± 5.7**
vs III	44.2 ± 7.3	45.1 ± 10.1	41.0 ± 5.3	26.3 ± 5.6
II vs II	20.9 ± 6.8	13.7 ± 9.0	34.6 ± 5.9	7.2 ± 4.9***
II vs III	38.1 ± 7.6	33.1 ± 9.4	32.8 ± 5.9	20.8 ± 6.2
III vs III	14.2 ± 3.5	57.4 ± 12.4***	15.5 ± 2.9	16.1 ± 4.2
	Transmembrane		Intracellular	
Comparison	p_{NC}	p_{NR}	p_{NC}	p_{NR}
I vs I	39.7 ± 3.4	23.2 ± 3.6***	42.1 ± 5.6	26.4 ± 6.5
I vs II	44.1 ± 3.1	31.3 ± 3.6**	44.6 ± 4.6	23.4 ± 5.1**
I vs III	43.0 ± 2.9	31.0 ± 3.5**	45.6 ± 4.6	27.4 ± 4.6**
II vs II	20.5 ± 2.8	8.6 ± 2.4**	29.6 ± 4.8	7.2 ± 4.0***
II vs III	31.8 ± 2.5	22.7 ± 3.2*	36.4 ± 4.2	24.6 ± 5.6
III vs III	19.4 ± 1.7	11.5 ± 1.7**	20.1 ± 2.7	17.5 ± 3.4

[a] Amino acids were categorized as nonpolar (A, F, I, L, M, P, V, W) or polar (all others). Any nonsynonymous difference leading to a change in polarity was counted as a radical difference (Hughes et al. 1990). Tests of significance of the difference between p_{NC} and p_{NR}: $*P < .05$; $**P < .01$; $***P < .001$

In certain comparisons among closely related subfamily III sequences, every amino acid difference in E2 involved a polarity difference (Fig. 2A). However, in comparisons between E2 domains of more distantly related olfactory proteins, the proportion of amino acid differences in this region that involved a polarity difference was lower (Fig. 2A). As a consequence, there was a negative relationship between the evolutionary distance between two olfactory proteins (as measured by the proportion amino acid difference between them) and the proportion of amino acid differences in E2 that involved a polarity difference (Fig. 2A). In E1, E3, and E4, exactly the opposite relationship held. Here, the proportion of amino acid differences that involved a polarity difference was positively related to the evolutionary distance between two proteins (Fig. 2B).

Discussion

The results suggest that there has been past directional selection to diversify polarity profile in the E2

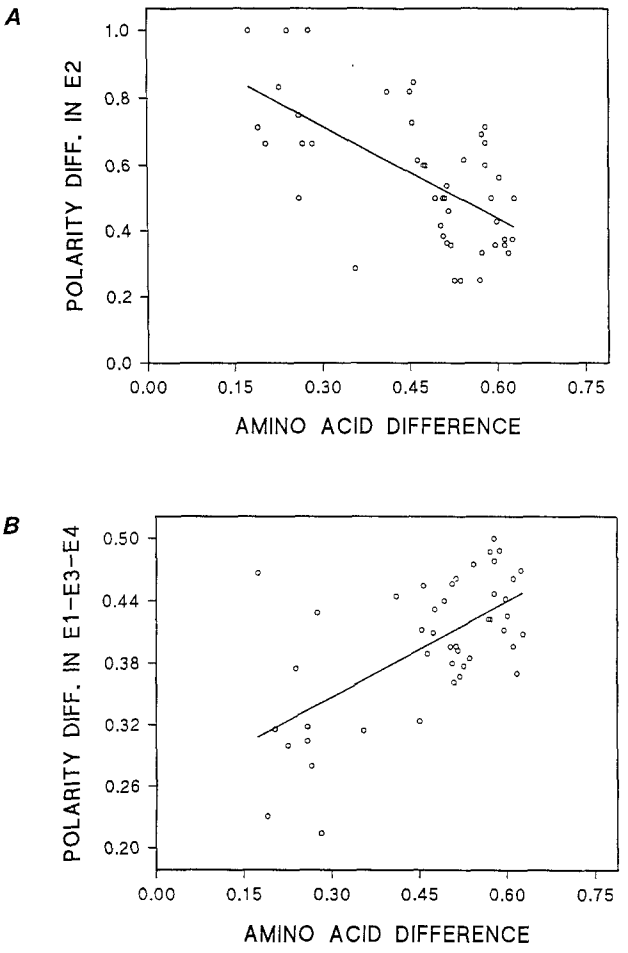

Fig. 2. The proportion of amino acid differences which involve a polarity change (*ordinate*) in E2 (**A**) and in E1, E3, and E4 (**B**) plotted against the mean proportion amino acid difference in the entire molecule (*abscissa*) for all pairwise comparisons among 10 rat olfactory proteins. Linear regression lines are drawn in each case. In A, r = −0.624; in B, r = 0.629.

domain of different members of the rat olfactory receptor gene family. The data on polarity differences in E2 of subfamily III members are consistent with the hypothesis that within this subfamily there has been a relatively rapid burst of positively selected amino acid replacements shortly after gene duplication (Goodman et al. 1975). On the other hand, comparisons between more distantly related genes suggest that, once functional divergence in E2 is achieved between two olfactory receptors, such directional selection no longer is present, and fixation of neutral mutants becomes the predominant mode of evolution. Such a process of evolution would explain the negative relationship between evolutionary distance and the proportion of amino acid changes involving a polarity change that is seen in the case of E2 (Fig. 2A). It would also explain why p_{NR} with respect to polarity in E2 greatly exceeds p_{NC} when the closely related members of subfamily III are compared, yet why p_{NC} seems to have become equal to p_{NR} when more distantly related sequences are compared. In E1, E3, and E4, however, the pattern is what would be expected under neutral evolution, with two proteins gradually accumulating more polarity differences as they diverge (Fig. 2B).

It would be difficult to explain the significant bias toward nonsynonymous differences causing a polarity change in E2 in comparisons among subfamily III genes on Kimura and Ohta's (1974) model whereby gene duplication is followed by a period in which one gene copy accumulates amino acid differences essentially without constraint. On this model, one might indeed expect that, shortly after gene duplication, certain nonsynonymous substitutions causing a polarity difference in E2 might occur by chance. But nonsynonymous changes in E2 not causing a polarity difference would be equally likely to occur.

This need not imply that Kimura and Ohta's model of evolution of new function after a period without function is not applicable in other cases. Indeed, this model seems well suited to explain certain cases reported in the literature. One example is the discovery that an interleukin 1 receptor antagonist is a member of the interleukin 1 gene family (Eisenberg et al. 1991). It is easy to imagine how a duplicate interleukin gene could have accumulated ordinarily deleterious mutations that would cause it to block the interleukin receptor without triggering an appropriate response and yet how, if appropriately regulated, such an antagonist molecule could play a useful role.

The mode of evolution seen in the case of the rat olfactory receptors may in fact be rather exceptional. It is perhaps significant that the best-documented cases of positive Darwinian selection at the molecular level have involved proteins playing a role in recognition of other molecules (Hughes and Nei 1988, 1989; Lee and Vacquier 1992) or in evading such recognition (Hughes 1991). Since the rat olfactory receptors have presumably diversified under selection favoring the ability to bind a wide variety of odorant molecules (Buck and Axel 1991), it is not surprising to see evidence of positive selection favoring diversification within this family.

The fact that the E2 domain is the focus of this selection suggests that E2 may play a role in odorant molecule recognition, perhaps as an odorant binding domain. Experimental evidence will be required to test this hypothesis. It is already known that I3 plays a role in interaction with the G protein (Hamm et al. 1988; Kobilka et al. 1988). Thus, the process of recombination or convergent evolution whereby I3 of F5 has come to resemble those of subfamily I (Fig. 1B) may have adaptive significance in that a novel receptor type has arisen which combines an E2 domain characteristic of subfamily

III with an I3 domain resembling subfamily I. Furthermore, the finding of adaptive evolution in the case of the rat olfactory receptors is of interest in that it provides evidence at the molecular level in support of the classical ethological hypothesis that genes affecting behavioral traits can be subject to adaptive evolution (Tinbergen 1951)

Acknowledgments. This research was supported by a grant from the National Institutes of Health to A.L.H. and by grants from the National Institutes of Health and the National Science Foundation to M. Nei.

References

Beets MGJ (1978) Odor and stimulant structure. In: Carterette C, Friedman MP (eds) Handbook of perception: vol. VIa, Tasting and smelling. Academic Press, New York, pp 245–255

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

Eisenberg SP, Brewer MT, Verderber E, Heimdal P, Brandhuber BJ, Thompson RC (1991) Interleukin 1 receptor antagonist is a member of the interleukin 1 gene family: evolution of a cytokine control mechanism. Proc Natl Acad Sci USA 88: 5232–5236

Goodman M, Moore GW, Matsuda G (1975) Darwinian evolution in the genealogy of hemoglobin. Nature 253:603–608

Hamm HE, Deretic D, Arendt A, Hargrove PA, Koenig B, Hofmann KP (1988) Site of G protein binding to rhodopsin mapping with synthetic peptides to the α subunit. Science 241:832–835

Hughes Al (1991) Circumsporozoite protein gene of malaria parasites (*Plasmodium* spp.): evidence for positive selection on immunogenic regions. Genetics 127:345–353

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–170

Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc Natl Acad Sci USA 86:958–962

Hughes AL, Ota T, Nei M (1990) Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility molecules. Mol Biol Evol 7:515–524

Kimura M, Ohta T (1974) On some principles governing molecular evolution. Proc Natl Acad Sci USA 71:2848–2852

Kobilka BK, Kobilka TS, Daniel K, Regan JW, Caron MG, Lefkowitz RJ (1988) Chimeric α_2-β_2-adrenergic receptors: delineation of domains involved in effector coupling and ligand binding specificity. Science 240:1310–1316

Lee Y-H, Vacquier V (1992) The divergence of species-specific abalone sperm lysins is promoted by positive Darwinian selection. Biol Bull 182:97–104

Li W-H (1983) Evolution of duplicate genes and pseudogenes. In: Nei M, Koehn RK (eds) Evolution of genes and proteins. Sinauer, Sunderland, MA, pp 14–37

Li W-H (1989) A statistical test of phylogenies estimated from sequence data. Mol Biol Evol 6:424–435

Mollon J (1991) Hue and the heptahelicals. Nature 351:696–697

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Nei M, Jin L (1989) Variances of the average numbers of nucleotide substitutions within and between populations. Mol Biol Evol 6:290–300

O'Dowd BF, Lefkowitz RJ, Caron MG (1989) Structure of the adrenergic and related receptors. Ann Rev Neurosci 12:67–83

Ohno S, Wolf U, Atkin NB (1968) Evolution from fish to mammals by gene duplication. Hereditas 59:169–187

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

Tinbergen N (1951) The study of instinct. Clarendon Press, Oxford, p 195

Wheeler JW (1977) Properties of compounds uses as chemical signals. In: Müller-Scwarze D, Mozell MM (eds) Chemical signals in vertebrates. Plenum Press, New York, pp 61–70

Received May 4, 1992/Accepted August 5, 1992