J Mol Evol (2005) 60:566–576
DOI: 10.1007/s00239-004-0172-y

JOURNAL OF MOLECULAR EVOLUTION

© Springer Science+Business Media, Inc. 2005

Adaptive Diversification of Vomeronasal Receptor 1 Genes in Rodents

Peng Shi,[1,2,3] Joseph P. Bielawski,[4,5] Hui Yang,[1,3] Ya-ping Zhang[1,2]

[1] Laboratory of Molecular Biology of Domestic Animals and Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China
[2] Laboratory for Conservation and Utilization of Bio-resources, Yunnan University, Kunming 650091, China
[3] The Graduate School, Chinese Academy of Sciences, Beijing 100039, China
[4] Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada
[5] Department of Mathematics & Statistics, Dalhousie University, Halifax, Nova Scotia, Canada

Received: 8 June 2004 / Accepted: 19 November 2004 [*Reviewing Editor*: Dr. Rasmus Nielsen]

Abstract. The vomeronasal receptor 1 (V1R) are believed to be pheromone receptors in rodents. Here we used computational methods to identify 95 and 62 new putative V1R genes from the draft rat and mouse genome sequence, respectively. The rat V1R repertoire consists of 11 subfamilies, 10 of which are shared with the mouse, while rat appears to lack the H and I subfamilies found in mouse and possesses one unique subfamily (M). The estimations of the relative divergence times suggest that many subfamilies originated after the split of rodents and primates. The analysis also reveals that these clusters underwent an expansion very close to the split of mouse and rat. In addition, maximum likelihood analysis showed that the nonsynonymous and synonymous rate ratio for most of these clusters was much higher than one, suggesting the role of positive selection in the diversification of these duplicated V1R genes. Because V1R are thought to mediate the process of signal transduction in response to pheromone detection, we speculate that the V1R genes have evolved under positive Darwinian selection to maintain the ability to discriminate between large and complex pheromonal mixtures.

Key words: Vomeronasal receptor — Positive selection — Gene duplication — Codon model

Correspondence to: Ya-ping Zhang; *email:* zhangyp@public.km.yn.cn or zhangypl@263.net.cn

Introduction

In mammals, the vomeronasal organ (VNO) of the olfactory system is a chemosensory organ specialized in the detection of pheromones, chemical signals that induce innate reproductive and social behaviors between the members of the same species (Halpern 1987; Wysocki and Lepri 1991; Keverne 1999). It is widely believed that pheromone perception is initiated by the interaction of chemical molecules with G protein-coupled receptors (GPCRs) in the VNO, and these candidate pheromone receptors are encoded by two distinct and complex superfamilies named V1R and V2R (Dulac and Axel 1995; Herrada and Dulac 1997; Matsunami and Buck 1997; Ryba and Tirindelli 1997). Expression of V1R genes is restricted to Gαi2-expressing sensory neurons of the apical zone of the VNO epithelium, whereas expression of V2R genes is exclusive to Gα0-expressing neurons of basal zone of the VNO epithelium (Berghard and Buck 1996; Jia and Halpern 1996). As in many other GPCR genes, V1Rs are characterized by an intronless coding region and a short extracellular N-terminus, while V2Rs have a large N-terminal domain (Dulac and Axel 1995; Herrada and Dulac 1997; Matsunami and Buck 1997; Ryba and Tirindelli 1997).

There are dramatic differences in the V1R repertoire among different mammalian lineages. The mouse genome contains about 300 V1R sequences classified into 12 gene families (Saito et al. 1998; Del Punta et al. 2000; Pantages and Dulac 2000;

Rodriguez et al. 2002). About 53% of these are believed to be pseudogenes (Rodriguez et al 2002). By contrast, the vast majority of V1R sequences in humans (~200) are believed to be pseudogenes, with only 5 known intact ORFs (Lane et al 2002; Zhang and Webb, 2003). The difference in V1R repertoire is consistent with morphological differences among lineages. The functional VNO was absent in adult catarrhine monkeys and humans while other nonprimate mammals, including rodents, possess a functional VNO (Keverne, 1999; Zhang and Webb, 2003). Not surprisingly, this observation led to the suggestion that such differences in V1R repertoires might reflect lineage-specific and species-specific differences in pheromone-related social and sexual behaviors (Lane et al. 2002; Rodriquez et al. 2002).

Lane et al. (2002) analyzed three major mouse V1R clusters (located on chromosome 6D, 6.56/57, and chromosome 13, respectively) and suggested that they arose via an expansion event associated with the divergence of mice and rats. Most recently, Lane et al. (2004) further analyzed two subfamilies in rats. Interestingly, both reports speculated that V1R expansion might contribute to differences in pheromone recognition, and thus played certain role in the process of rodent speciation. However, their analyses of V1R from either the mouse or the two subfamilies alone provide only a limited evolutionary perspective of this important family. We expected that a comparative analysis of this gene family from a closely related species would be a powerful way to further illustrate the molecular evolution of this family, especially the role of adaptive evolution of these genes in rodents. Therefore, we used the recent releases of the rat genome sequence (Rat Genome Sequencing Consortium 2004) to identify the V1R genes of rat and conducted an analysis of this gene family evolution in mice and rats.

Methods

Data Mining and Sequence Quality Control

Previously identified V1R genes from mouse and human were used as queries to screen the rat genome sequences of the June 2003 assembly (http://genome.ucsc.edu) by using the program BLASTN or TBLASTN (Altschul et al. 1997).

Initially, we included all previously reported mouse genomic sequences in our mouse dataset. However, when we compared the mouse V1R genes previously identified from the Celera database by Rodriguez et al. (2002) with the latest assembly of draft mouse genome (October 2003 assembly; http://genome.ucsc.edu), we found that some sequences from the new assembly contained frame shifts as a result of the insertion of one or two nucleotides, leading to stop codon interruption. For example, the new sequence of V1RC19 is different from the old one in having an additional nucleotide (G) at position 872. Furthermore, some genes previously regarded as distinct actually represent only one genomic position in the new assembly. For instance, V1RB11 is the same as the V1RB4 in the later assembly. The most plausible explanation is that these discrepancies are artifacts of the data handling procedures that have been resolved in the later assemblies. In fact, the V1RB11 gene is no longer present in the update Celera database (Del Punta et al. 2002). We excluded 13 such problematic sequences from this study (V1RB5, V1RB6, V1RB11, V1RC1, V1RC4, V1RC13, V1RC19, V1RC22, V1RH1, V1RI8, V1RD1, V1RD2, V1RD18), although some real functional genes might have also been removed under this criterion.

Here we adopted the classification and nomenclature proposed by Rodriguez et al (2002). Each V1R gene is denoted by one additional letter followed by one number. The letter represents the subfamily to which the gene belongs and the number designates the order of discovery of the genes in that subfamily. For instance, the notation rV1RA1 indicates that the sequence is from the rat (r) and the first member identified as belonging to the A subfamily. The classification of subfamily was based on the phylogenetic relationships and sequence homology among sequences using the criteria proposed by Rodriguez et al. (2002).

Evolutionary Analyses

Deduced V1R amino acid sequences were initially aligned by CLUSTAL W (Thompson et al. 1994) and then the alignment was refined by manual adjustment. The nucleotide sequences were aligned according to the aligned amino acid sequence and used in subsequent tree reconstruction. The alignments are available from the authors on request. The mouse and human T2R genes, members of the bitter taste receptor gene family, were used as outgroups. The phylogenetic relationships of the V1R gene sequences were reconstructed by using the neighbor-joining method (Saitou and Nei 1987) under a variety of models (uncorrected P-distance, K2P [Kimura 1980], JC69 [Jukes-Cantor 1969], and HKY [Hasegawa et al. 1985]) with program MEGA (Kumar et al. 2001). In addition, a more sophisticated clustering procedure, BioNJ, under various models of evolution with parameters, including proportions of invariable sites (I) and gamma distribution (γ), were also used by Phyml program (Guindon and Gascuel 2003) to construct the phylogenetic trees. Since all these analyses gave similar results, only NJ trees under JC69 are presented here. The reliability of the trees obtained was evaluated by the bootstrap method (Felsenstein 1985) with 1000 pseudo-replications. Sawyer's (1989) method was used to examine gene conversion among paralogous genes. We used the two-cluster test of Takezaki et al. (1995) to examine molecular clock hypothesis. The linearized-tree method of Takezaki et al. (1995) was employed to estimate the relative divergence times of each V1R subfamily.

To examine the pattern of nucleotide substitution, the numbers of synonymous substitutions per synonymous site (d_S) and the number of nonsynonymous substitutions per nonsynonymous site (d_N) were estimated by maximum likelihood method using the codeml program of the PAML package (Yang 1997). Note that those V1R genes involved in gene conversions were excluded from this analysis, because such sequences can increase the rate of false positives when testing the positive selection (Anisimova et al. 2001). Two kinds of codon-substitution models were employed in this paper. First, we used "site-specific" models to analyze each of the subfamilies or clusters whose origins were potentially associated with the divergence of the mouse and rat lineages. Here we employed the recommended models (M0, M3, M7, and M8) (Anisimova et al. 2001, 2002). M0 is the "one-ratio" model, having one ω (d_N/d_S) parameter for all sites in the data. M3 is the "discrete" model, having a discrete number of site categories, each with an independent ω parameter. In this case we used M3 with $k = 3$ site categories. M7 ($k = 10$) is the "beta" model, as it uses a beta distribution to model variable selection pressure among sites. M8 ($k = 11$) is the "*beta & ω*" model, an extension of M7 that allows for an extra ω parameter that can be >1. Likelihood ratio tests (LRTs) were conducted to examine whether selection pressure varied among sites and a portion of sites was under positive selection.

Second, we employed "branch-site" models of Yang and Nielsen (2002) to test for positive selection at sites along prespecified sets of branches. We conducted this type of analysis on subfamily A because this family contained large samples of mouse and rat sequences, allowing tests for episodes of positive Darwinian selection following both gene duplication events and speciation events within the same clade.

The branch-site models assume that selection pressure (ω) varies among sites and at a subset of these sites, selection pressure changes in a specified set of branches called the foreground (FG) branches. All other branches are referred to as background (BG) branches. Yang and Nielsen (2002) implemented two versions, called Model A and Model B. We choose to employ Model B, given Model B imposes less restrictions on the data than Model A (Yang and Nielsen 2002).

LRTs were used to test for (i) divergent selection pressure at a fraction of sites in specific branches of the gene tree and (ii) a ω parameter greater than 1 at such sites in those branches of the gene tree. The first LRT compares a null model that permits variation in ω among sites but not branches (M3; $k = 2$) with a model that permits variable ω among sites and between FG and BG branches (Model B). Note that LRT only tests for divergent selection pressure in the FG branch. With regard to identifying positively selected sites, the error rate can be excessively large when a fraction of sites is evolving under strict neutrality, e.g., $\omega = 1$ (Hayden et al. 2001; Anisimova et al. 2002; Zhang et al. 2004), which is not distinguished by this LRT. To avoid false positives in such a case, we implemented a second LRT that compares a rigid null model where the FG branches were constrained to neutral evolution (Model B; $\omega = 1$) with a model where positive selection is permitted in the FG branches (Model B; $\omega \geq 1$) (Bielawski et al. 2004). Only in cases where the second LRT was significant, did we use an empirical Bayesian approach to predict which sites were most likely to have evolved under positive selection (Nielsen and Yang 1998).

Results

Identification of Rat and Mouse V1R Functional Genes

The lack of introns in V1R genes allowed us to use a BLAST-based approach to identify a large number of presumably functional V1R genes in the rat genome sequence. For the convenience of comparison with previously identified mouse sequences, we searched the rat draft genome sequences for new V1R genes using the same stringent criteria as those by Rodriguez et al. (2002). Only those ORFs longer than 810 bp or those containing seven intact transmembrane domains were considered as putative V1R genes. A total of 102 putative rat V1R functional genes were ultimately identified, including seven genes that had been reported previously (Dulac and Axel 1995) (Table 1). We also searched the latest draft mouse genome sequence and identified 62 putative V1R genes that were missed in previous reports (Saito et al. 1998; Del Punta et al. 2000; Pantages and Dulac 2000; Rodriguez et al. 2002) (Table 1). Thus, the final set of mouse putative functional genes used in this study was comprised of 62 newly identified genes and 125 previously reported genes. Although our findings, together with the 13 problematic sequences listed in the methods section indicate a real potential for both false-negative and false-positive annotation errors, the number of each V1R subfamily in mouse and rat (Table 1) suggests that the newly discovered rat V1R genes may make up a large proportion of the members in most subfamilies.

Table 1. Numbers of genes of each V1R subfamily in mouse and rat

Subfamily	Mouse	Rat
A	10	8 (5)
B	9 (1)	6 (2)
C	32 (4)	23 (23)
D	56 (39)	8 (8)
E	16 (3)	22 (22)
F	5	8 (8)
G	21 (9)	13 (13)
H	23 (3)	0
I	12 (3)	0
J	1	4 (4)
K	1	2 (2)
L	1	7 (7)
M	0	1 (1)
Total	187 (62)	102 (95)

Note. The number of V1R genes after removing the questionable sequences. The number of newly identified genes in each subfamily is given in parentheses.

The mouse V1R repertoire was comprised of 12 subfamilies. The rat V1R repertoire consisted of 11 subfamilies, 10 of which were shared with the mouse. It lacked the H and I subfamilies found in mouse and possessed one unique subfamily (M), which was comprised of only one gene sequence. The new M gene sequence (rV1M1) shows all features typical of the V1R family, including the fourteen amino acid residues conserved among 95% of mouse V1Rs and a potential N-linked glycosylation site in extracellular loop II conserved in all mouse V1Rs (Rodriguez and Mombaerts 2002). In addition, we found that a pseudogene sequence in mouse genome has 83% similarity with rV1RM1 (data not shown). Bootstrap analysis also indicated strong support for a sister relationship between the M sequence and the clade comprised of the A and B subfamilies (Fig. 1a). Taken together, these data strongly support that the M sequence is a new component of V1R family.

In mouse, the H and I subfamilies are distributed in a cluster on chromosome 13, which is homologous to rat chromosome 17. Surprisingly, no orthologs of the mouse H and I subfamilies were identified in our initial search of the rat genome sequences. However, when mouse V1RI genes were used as queries, a sequence segment with very high similarity was identified. The ORF of this segment contained a premature stop codon. Thus, it appears that the only rat se-

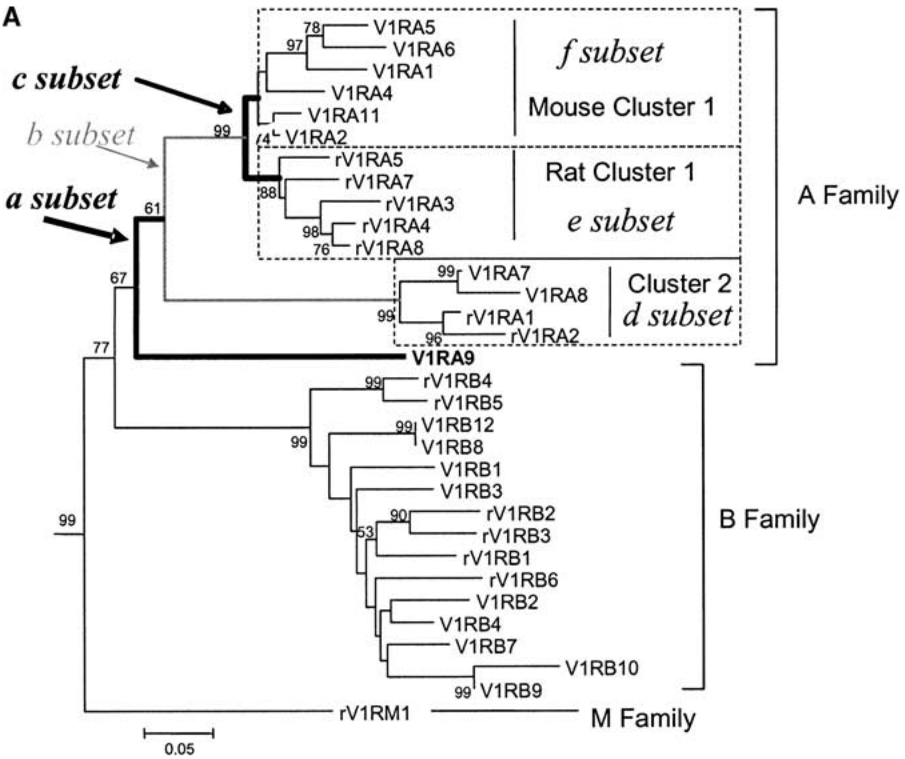

Fig. 1. Neighbor-joining tree of 289 putatively functional V1R genes from the rat and mouse. After the removal of gaps, a total of 552 nucleotide sites is used in reconstructing the tree. Genetic distance was based on the Jukes–Cantor model. **A** Subtree of the A, B, and M subfamilies. The V1R genes and branches in the A subfamily were classified into six subsets. Each of them was separately specified as foreground (FG) in the branch-site model of Yang and Nielsen (2002). Subset *a* was comprised of the basal lineages of the A subfamily, exclusive of clusters 1 and 2. Subset *b* was comprised of the branches that immediately postdate the duplication event that gave rise to clusters 1 and 2 of the A subfamily. Subset *c* was comprised of the branches that immediately postdate the mouse and rat speciation event resolved within cluster 1. Subset *d* included all branches of cluster 2. Subset *e* consisted of all branches of the rat clade of cluster 1. Subset *f* consisted of all branches of the mouse clade of cluster 1. Percentage bootstrap values (≥50) are shown on interior branches. **B** Topology of whole V1R gene families. The T2R genes are used as the outgroup. The bootstrap values of each branch separating the various subfamilies are indicated. A color figure is available as online Supplementary Material.

quence segment that is potentially orthologous to the I subfamily of mouse is likely to be a pseudogene. In the mouse genome, the H and I subfamily clusters are flanked by the histone 1 gene (NM_178198) and the Abt1 gene (NM_013924). Coincidently, the rat sequence segment identified above is flanked by the same two genes. Thus, it appears that at least one sequence ancestral to the I subfamilies existed before the divergence of the mouse and rat lineages. When mouse H subfamily genes were used as queries, a sequence of high similarity was identified on rat chromosome 15 rather than 17. Hence, the origin of H might predate the divergence of mouse and rat as well.

Phylogenetic Relationships of Functional V1R Family from Mice and Rats

Figure 1b shows a phylogenetic tree inferred from 289 putatively functional V1R genes of the rat and mouse. As a representative family, the subtree of families A, B, and M is presented in Fig. 1a. Other subtrees of each family are separately provided online as supplement 1. Phylogenetic analysis showed that the newly identified rat genes represent 12 evolutionary lineages with strong supports. Moreover, these data reveals that mouse and rat sequences do not form two separate clusters in each subfamily. They appear intermingled with each other. This result suggests that many gene duplication events predated the separation of mice and rats. By contrast, some V1R genes formed species-specific clades. For example, within the A subfamily (Fig. 1a) mouse V1R genes, V1RA1, V1RA11, V1RA2, V1RA3, V1RA4, V1RA5, and V1RA6 were grouped together, and the rat V1R genes rV1RA3, rV1RA4, rV1RA5, rV1RA6, rV1RA7, and rV1RA8 formed another separate clade. These species-specific genes may be the result of gene conversions and/or postspeciation duplications. Gene conversions among paralogous genes were detected by Sawyer's (1989) test for both the rat and the mouse. Among the 289 mouse and rat V1R genes, only 25 gene conversion events were detected, suggesting that gene conversions may not have

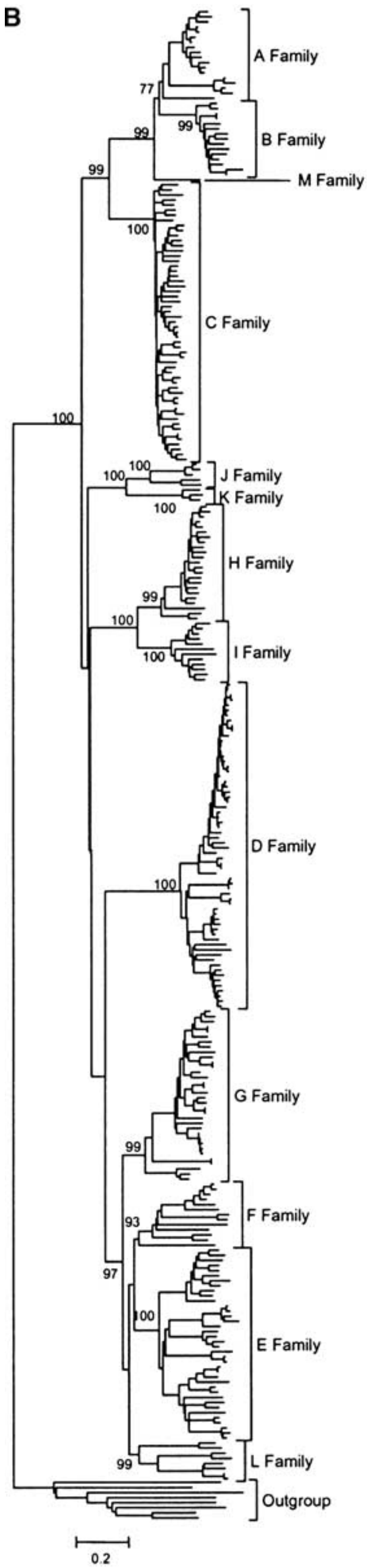

Fig. 1. (Continued).

played a major role in the evolution of V1Rs; being infrequent and acting locally. Thus, most species-specific genes may be formed by postspeciation duplications and might have species-specific functions or perhaps could relate to reinforcement of speciation, as has been suggested by Lane et al. (2002).

Relative Times of Expansion of Different Gene Families and Clusters

Since almost all V1R-like sequences found in humans seem to be pseudogenes (Pantages and Dulac 2000; Rodriguez et al. 2000; Giorgi et al. 2000; Lane et al. 2002), it is interesting to estimate the relative divergence times of the functional genes found in rodents. The two-cluster test of Takezaki et al. (1995) revealed that the molecular clock hypothesis could not be rejected in our all cases. A linearized tree for these data is shown in Fig. 2. It is clear from the topology that this family emerged between 80 million and 100 million years ago, during the proposed radiation of placental mammals (80–110 MYA; Springer et al. 2003), and many subfamilies originated after the split of rodents and primates (about 75 MYA; Springer et al. 2003), indicating that they are rodent-specific genes. Figure 2 also reveals a wide rage of divergence dates for the subfamilies, yet most clusters within different subfamilies underwent an expansion during a relatively short period of time.

It can be very useful to estimate dates for such events, however, estimating molecular dates for rodents is complicated by the need to employ calibration points from other lineages where there is significant rate heterogeneity (Adkins et al. 2001, 2003). As rodent V1R genes are evolving in a clock-like manner, it is possible to calibrate a local time scale of their evolution. Although in this case the relative ages obtained may be correct, the specific estimates of dates could be biased. We inferred such a local time scale by assuming the mean nucleotide substitution rate of 5×10^{-9}per year per site for rodent evolution (O'Huigin and Li 1992), which is extremely similar to the substitution rate estimated from the entire genomes (Rat Genome Sequencing Consortium 2004). Based on this calibration, many subfamilies underwent an expansion between 12 and 20 MYA (Fig. 2), very close to the mouse–rat split time. This estimation was consistent with that of Lane et al. (2002), which was obtained by analyzing mouse V1R gene-block duplications.

Adaptive Diversification After Gene Duplications in the V1R Family

Why did so many genes from different subfamilies undergo such a substantial expansion over a short time period? To help answer this question, we attempted to resolve the mode and tempo of selection pressure acting on the V1R gene family. First we

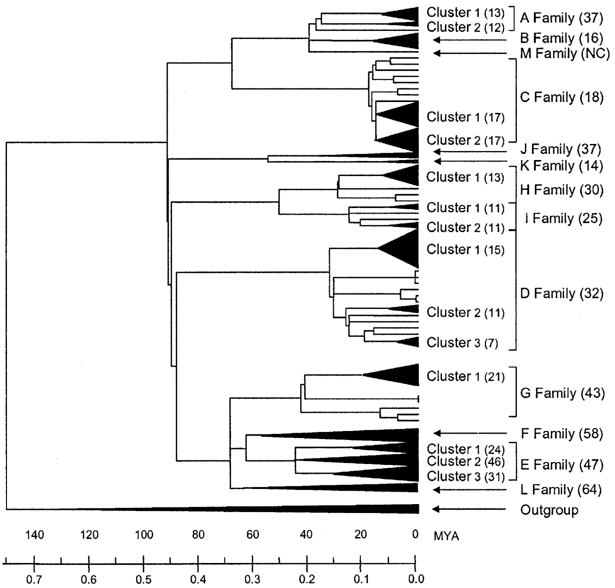

Fig. 2. Linearized tree for the V1R superfamily. The divergent times and nucleotide distances between different subfamilies are marked on the scale below the tree. The detailed divergent times of each cluster and subfamily are given in parentheses. The sizes of black triangles correspond to the numbers of genes within the clade.

focused on the clusters which underwent an expansion between 12 and 20 MYA (marked in Fig. 2) and separately estimated the distribution of selection pressures (ω) among sites in each cluster using the "site-specific" models. Estimates of the parameter values of the ω distribution under M3 (data not shown) and M8 (Table 2) indicate that a fraction of sites is under positive selection in each of the subfamilies tested. LRTs of M0 and M3 were significant and confirmed that selection pressure varied among sites (Table 2). Moreover, the LRTs of M7 and M8 were significant in all cases, indicating that some variation in selection pressure was due to the evolution of a subset of sites by positive Darwinian selection (Table 2).

To more precisely resolve the mode of adaptive evolution, we performed a more detailed analysis of the genes in the A subfamily using "branch-site" models. As indicated in Figure 2a, six nonintersecting subsets (*a*–*f*) of the A subfamily were specified as the foreground (FG) branches and the complement was specified as the background (BG) in each analysis. Parameter estimates under Model B indicated that a fraction of sites (1–9%) is evolving by positive selection in four of the six subsets of branches (Table 3). LRTs were performed to test the improvement obtained by allowing heterogeneous selection pressure between the BG and FG branches (Model B). Note that in this case the null model only allowed variable selection pressure among sites (M2; $k = 2$). Results (Table 3) indicated that a significant improvement was achieved by allowing heterogeneous selection pressures in branch subsets *a*, *d*, and *e*. Subset *a* consisted of the basal lineages of the A subfamily, subset *d* consisted of all branches in cluster 2, and subset *e* consisted of all branches of the rat clade of cluster 1 (Fig. 1a).

The LRT above is a test of heterogeneous selection pressure between the FG and BG branches. We also wanted to test whether such heterogeneity was caused by the positive selection in the FG branches. Hence, we conducted an LRT that compared a null hypothesis having sites in the FG branches constrained to a rigid neutral model (Model B with $\omega_2 = 1$) to an alternative hypothesis that permitted those branches to have values consistent with positive selection (Model B with $\omega_2 \geq 1$). Table S1 contains the parameter estimates and likelihood scores under the null and alternative hypotheses. In each case (subsets *a, d*, and *e)*, the result of this LRT was significant (Table S2), indicating that the estimates of ω_2 in those cases were significantly larger than 1. Thus, the LRTs are consistent with positive selective pressure driving functional divergence V1R genes.

Since the parameter estimates were consistent with positive selection, and significant LRTs were obtained, it was desirable to infer which sites are under positive selection in each subset. In this case we wanted to know if the target of positive selection had changed over evolutionary time in the A subfamily. We examined the posterior probabilities for site classes under Model B. At the $P > 50\%$ level, 10 sites were identified as potential targets of positive selection in subset *a*, and 11 sites in subset *d* (Table 4). None of the sites identified in subset *d* matched those in subset *a*, suggesting a dramatic shift in the target of positive selection. When 3D structures of this protein are available in the future, we can investigate whether these putative positively selected sites are located in the ligand binding regions.

Sampling errors are known to be high for ML parameter estimation under complex codon models (Anisimova et al. 2001, 2002). Our above results were obtained under the F3×4 model of codon frequencies, which uses the nucleotide frequencies at the three positions of the codon to compute the equilibrium codon frequencies (Goldman and Yang 1994). Because the reliability of Bayesian site identification depends on the reliability of the ML parameter estimates, it is important to assess the robustness of the results under different model assumptions. Therefore, we also estimated the parameters of Model B under a different model (F61) that uses empirical estimates of the 61 codon frequencies. Results of parameter estimation and LRTs were in good agreement under both models (data not shown), with Bayesian site identification being identical for subsets *d*, *e*, and *f* and only a few differences observed for subset *a* (Table 4).

Table 2. LRT of positive selection for the subfamilies and clusters expanded during the mouse–rat split

			$2\Delta l$			
Cluster	n	Lc	M3 vs. M0	M8 vs. M7	Parameters estimated under M8	Positively selected sites
A subfamily	16	813	65.88***	13.72**	$\mathbf{f_1 = 0.00609}$, ω = 5.53572, f_0 = 0.99391, p = 1.08815, q = 1.15069	60M
B subfamily	14	870	78.94***	28.82***	$\mathbf{f_1 = 0.03509}$, $\boldsymbol{\omega} = \mathbf{4.16258}$, f_0 = 0.96491, p = 0.84814, q = 0.58204	75Q 150L 185C 226I 251N 255H
C subfamily cluster 1	16	888	131.24***	25.6***	$\mathbf{f_1 = 0.20140}$, $\boldsymbol{\omega} = \mathbf{1.58006}$ f_0 = 0.79860, p = 2.23719, q = 4.78622	62K 71G 117H 141R 151V 155N 196T 290I 295Q 296K
C subfamily cluster 2	20	876	88.36***	8.08*	$\mathbf{f_1 = 0.19380}$, $\boldsymbol{\omega} = \mathbf{1.18918}$ f_0 = 0.80620, p = 1.55156, q = 2.90533	106V 165Y 191T
D subfamily cluster 1	37	906	202.98***	60.92***	$\mathbf{f_1 = 0.08075}$, $\boldsymbol{\omega} = \mathbf{3.18749}$ f_0 = 0.91925, p = 0.52968, q = 0.45891	8L 134T 224D 225A 235H 265R
D subfamily cluster 2	7	912	50.5***	15.42***	$\mathbf{f_1 = 0.03590}$, $\boldsymbol{\omega} = \mathbf{6.15484}$, f_0 = 0.96410, p = 0.09824, q = 0.10950	259L
D subfamily cluster 3	8	900	74.3***	9.22**	$\mathbf{f_1 = 0.12741}$, $\boldsymbol{\omega} = \mathbf{2.62462}$, f_0 = 0.87259, p = 0.06912, q = 0.09820	164S
H subfamily cluster 1	20	891	324.24***	60.64***	$\mathbf{f_1 = 0.10778}$, $\boldsymbol{\omega} = \mathbf{2.77467}$, f_0 = 0.89229, p = 0.30760, q = 0.35257	8K 132S 153S 163D 260R 263N 269R 292V
I subfamily cluster 1	5	882	56.6***	10.14***	f_1 = 0.17068, ω = 2.23142, f_0 = 0.82932, p = 1.07561, q = 3. 29489	6H 14T 26V 28H 29I 53I 58S 61I 81T 82V 270T 291V
I subfamily cluster 2	5	894	87.34***	17.86***	$\mathbf{f_1 = 0.04295}$, ω = 5.97888, f_0 = 0.95705, p = 0.05494, q = 0.09937	210V 287W

Note. *Significant at 5% level.**Significant at 1%. ***Significant at 0.1%.
[a]Sites with posterior probabilities >95% under M8 are listed.

Table 3. Parameter estimates and likelihood scores for the A subfamily genes

Model	p	Estimates of ω parameters	Positive selection	l	$2\Delta l$, M3 ($k = 2$) vs. Model B
M0: One-ratio	1	$\omega = 0.50$	None	−6601.02	
Site-specific models					
M3: Discrete ($k = 2$)	3	$\omega_0 = 0.22, f_0 = 0.46$ $\omega_1 = 0.80, (f_1 = 0.54)$	None	−6566.70	
Branch-site models					
Model B: subset *a*	5	$\omega_0 = 0.22, f_0 = 0.43,$ $\omega_1 = 0.78, f_1 = 0.49,$ $\boldsymbol{\omega_2 = 7.15}, (f_1 = 0.09)$	10 sites[a]	−6561.91	**9.588****
Model B: subset *b*	5	$\omega_0 = 0.22, f_0 = 0.43,$ $\omega_1 = 0.81, f_1 = 0.50,$ $\omega_2 = 0.00001, (f_1 = 0.07)$	None	−6566.54	0.308
Model B: subset *c*	5	$\omega_0 = 0.22, f_0 = 0.36,$ $\omega_1 = 0.81, f_1 = 0.43,$ $\omega_2 = 0.00001, (f_1 = 0.21)$	None	−6566.47	0.462
Model B: subset *d*	5	$\omega_0 = 0.22, f_0 = 0.46,$ $\omega_1 = 0.80, f_1 = 0.49,$ $\boldsymbol{\omega_2 = 9.37}, (f_1 = 0.05)$	11 sites[b]	−6553.45	**26.50*****
Model B: subset *e*	5	$\omega_0 = 0.19, f_0 = 0.41,$ $\omega_1 = 0.75, f_1 = 0.58,$ $\boldsymbol{\omega_2 = 7.89}, (f_1 = 0.01)$	1 site[c]	−6562.87	**7.652***
Model B: subset *f*	5	$\omega_0 = 0.20, f_0 = 0.41,$ $\omega_0 = 0.75, f_1 = 0.58,$ $\boldsymbol{\omega_2 = 5.85}, (f_1 = 0.01)$	2 sites[d]	−6564.77	3.855

Note. Equilibrium codon frequencies were obtained under the F3×4 model. Frequency parameters, *f*, shown in parentheses were obtained by subtraction. *p* is the number of free parameters in the ω distribution. Significant LRTs are in boldface.

*Significant at 5% level. **Significant at 1%. ***Significant at 0.1%.

[a] 4 N 0.5637, 61 A 0.8906, 128 A 0.7248, 144 L 0.5849, 155 N 0.5997, 162 0.7175, 167 L 0.7017, **234 M 0.9766**, 252 S 0.5665, 256 F 0.5067.

[b] 119 R 0.7889, 212 H 0.5719, **241 L 0.9568**, 258 H 0.5110, **259 I 0.9125**, 260 Y 0.7256, **262 T 0.9798, 263 V 0.9403, 266 F 0.9955**, 267 V 0.6396, **270 S 0.9455**.

[c] **60 M 0.9937**.

[d] **60 M 0.9167**, 154 T 0.7547.

Table 4. Predicted positive selection sites for subsets of the V1R phylogeny inferred under Model B and two different methods of computing codon frequencies

Subset *a*		Subset *d*		Subset *e*		Subset *f*	
F3×4	F61	F3×4	F61	F3×4	F61	F3×4	F61
4 N 0.5637	4 N 0.6113	119 R 0.7889	119 R 0.7601	**60 M 0.9937**	**60 M 0.9976**	**60 M 0.9167**	**60 M 0.9648**
61 A 0.8906	9 T 0.5141	212 H 0.5719	212 H 0.5580			154 T 0.7547	154 T 0.6687
128 A 0.7248	61 A 0.8921	**241 L 0.9568**	**241 L 0.9709**				
144 L 0.5849	128 A 0.6969	258 H 0.5110	258 H 0.5160				
155 N 0.5997	144 L 0.5919	**259 I 0.9125**	**259 I 0.9495**				
162 Q 0.7175	149 T 0.6632	260 Y 0.7256	260 Y 0.5977				
167 L 0.7017	155 N 0.5745	**262 T 0.9798**	**262 T 0.9888**				
234 M 0.9766	162 Q 0.6978	**263 V 0.9403**	**263V 0.9235**				
252 S 0.5665	167 L 0.7074	**266 F 0.9955**	266 F 0.9980				
256 F 0.5067	**234 M 0.9664**	267 V 0.6396	267V 0.8015				
	252 S 0.5206	**270 S 0.9455**	**270 S 0.9434**				

Note. Branches of the V1R phylogeny corresponding to each subset are depicted in Fig. 1a. The sites were indexed by their position in the alignment and by the amino acid at the site in V1RA1 (AF291481). Positive selection sites were identified using the empirical Bayes approach with posterior probability $p \geq 50\%$, with those ≥90% in boldface. F3×4 indicates that codon frequencies were computed from the frequencies of nucleotides at the three potions of the codon. F61 indicates that the empirical estimates of the 61 codon frequencies were used.

Discussion

In this paper, we took advantage of the draft rat genome sequence to further study V1R gene family evolution in rodents. Using computational methods we identified 102 and 62 potentially functional V1R genes from the rat and mouse genome, respectively. Together with previously identified rodent V1R genes (Saito et al. 1998; Del Punta et al. 2000; Pantages and Dulac 2000; Rodriguez et al. 2002), we conducted a comparative analysis in rodents. A new subfamily of V1R genes (M) that appears to be unique to the rat lineage and two subfamilies apparently unique to mouse (H and I) were identified. In addition, most of the V1R genes within different rodent subfamilies exhibit species-specific gene clusters, suggesting the presence of "species (lineage)-specific" vomeronasal pheromone receptors. Groups of these receptors may have evolved separately in different species to deal with the specific types of chemicals they encounter. Interestingly, species-specific clusters are also observed in the rodent T2R family, where it is thought that they could reflect the adaptive evolution of species-specific bitter taste receptors (Shi et al. 2003).

Relative divergence dates for rodent V1R subfamilies were variable, but no date was earlier than the divergence date of primates and rodents. Hence, V1R diversity in rodents reflects the interplay of evolutionary forces originating within the rodent lineage. Interestingly, if we assume the rate estimated by O'hUigin and Li (1992), a large number of clusters within subfamilies occur close to the divergence time of mouse and rat, with additional expansions having occurred recently within both mouse and rat (as recently as 9 MYA).

Lane et al. (2002) noted that expansion of three V1R loci sampled from the mouse genome appeared to coincide with the divergence of mouse and rat. A remarkable finding of this study was that expansion of most of the V1R gene clusters within different rodent subfamilies seemed to occur very close together in time. This suggested a gene family–wide process of expansion. To determine if positive Darwinian selection could have played a role during the evolution of these clusters, we measured selection pressure in these clusters using the d_N/d_S ratio (Yang and Bielawski 2000). We detected a small fraction of sites evolving under positive selection in each cluster. Although evolution by positive selection is not uncommon in sensory receptor gene families (Ngai et al. 1993; Hughes and Hughes 1993; Shi et al., 2003), our findings, and those of Lane et al. (2002), are unusual in that they raise the possibility that positive selection might have been related to a species level divergence event.

We conducted a further analysis of selection pressure within the A subfamily because the mouse–rat divergence, as well as a gene duplication event, was clearly resolved within this subfamily. Using codon models that allowed selection pressure to vary both among sites and branches, we discovered an episodic mode of positive selection. There was a strong signal for positive selection in some parts of the A subfamily gene tree, and none in other parts. Moreover, the sites targeted by positive selection differed among the different parts of the tree. Surprisingly, no evidence for positive selection was found in the branches associated with the divergence of mouse and rat or in the branches associated with the divergence of clusters 1 and 2 within the A subfamily. Rather, the general pattern was one of adaptive evolution during the expansion of clusters (mouse cluster 1, rat cluster 1, and cluster 2) that followed the divergence of mouse and rat.

The association of the V1R family with detection of pheromones affecting social and sexual behaviors is intriguing, as genes that mediate sexual reproduction are often more divergent than nonproductive genes (e.g., reviewed by Swanson and Vacquier 2002). Although the composition and variability of pheromonal mixtures are largely unknown among mammals, they have been widely used to detect the social and reproductive status of potential mates (Halpern 1987; Wysocki and Lepri 1991). A mixture of pheromones evolving by a rapid neutral process could create strong selection pressure for a recognition system such as V1R to quickly adapt to pheromonal changes. An analogy can be made with the rapid neutral evolution in abalone VERL that creates strong selection pressure for rapid adaptive change in lysin (Metz et al. 1998). In the case of abalone, neutral mutations accumulate differently in VERL in different populations, and lysin in different populations must adapt in different ways in order to maintain an efficient VERL–lysin interaction (Metz et al. 1998). In the case of rodents, if pheromone mixtures evolved rapidly and accumulated neutral changes in isolated populations, then recognition systems such as V1R could be subject to different selective pressures in different populations. This could lead to lineage-specific incompatibility of pheromone recognition systems. If a large and complex V1R repertoire is required to discriminate between pheromone blends, we might then expect positive selection pressure to have acted across many members of the V1R family. Interestingly, many members of the V1R family were subject to adaptive evolution, as revealed by our result

Genes that mediate reproduction can be the direct target of sexual selection (e.g., reviewed by Swanson and Vacquier 2002), in which case the above model would not be correct. For instance, if pheromone-mediated mate selection is important in rodents, the characteristics of the pheromonal mixtures might themselves be under intense sexual selection. In this case adaptive co-evolution of the associated pheromonal recognition system, such as V1R, would be driven by the strength of sexual selection on the pheromonal mixture. If sexual selection played a prominent role in the mouse–rat speciation process, we might expect to see a strong signal for adaptive evolution in reproductive genes along the branches most closely associated with the speciation event. In this study, we were unable to detect such a signal in the A subfamily of V1R.

We have argued for the possibility that evolution of pheromone blends in isolated populations of rodents could have generated selective pressure for adaptive co-evolution of the V1R repertoire. Given enough time, such an evolutionary process could lead to reproductive isolation through divergence in pheromone-based mate selection systems. Pheromone and receptor co-evolution has been suggested to play a role in the speciation process of some insects (e.g., Roelofs and Rooney 2003). However, we observed that adaptive evolution only occurred in V1R clusters that expanded following the mouse–rat speciation event, and not along those branches associated with the speciation event itself. Based on these findings, we suggest that co-evolution of pheromone blends and a V1R-based recognition system might have played an important role in reinforcing barriers to fertilization among species of rodents, rather than being directly involved in the speciation event. Such issues cannot be resolved for rodents without further studies along two different fronts. First, studies are needed to determine the importance of pheromone blends in distinguishing between one's own species and a closely related species of rodent. Second, the molecular evolution of the V1R family needs to be investigated within additional and more closely related species of rodents.

Acknowledgments. We thank Katherine A. Dunn and two anonymous reviewers for constructive comments. This work was supported by grants from the Chinese Academy of Sciences (KSCX2-1-05), the National Natural Science Foundation of China (30430110 and 30021004), and the Program for Key International S & T Cooperation Project of P.R. China (2001CB711103) to Y.P.Z, a startup fund from Dalhousie Unievrsity to J.P.B., and a grant from the Genome Atlantic Centre of Genome Canada to J.P.B.

References

Adkins RM, Gelke EL, Rowe D, Honeycutt RL (2001) Molecular phylogeny and divergence time estimates for major rodent groups: evidence from multiple genes. Mol Biol Evol 18:777–791

Adkins RM, Walton AH, Honeycutt RL (2003) Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes. Mol Phylogenet Evol 26:409–420

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 25:3389–402

Anisimova M, Bielawski JP, Yang Z (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol 18:1585–1592

Anisimova M, Bielawski JP, Yang Z (2002) Accuracy and power of bayes prediction of amino acid sites under positive selection. Mol Biol Evol 19:950–958

Berghard A, Buck LB (1996) Sensory transduction in vomeronasal neurons: evidence for G alpha o, G alpha i2, and adenylyl cyclase II as major components of a pheromone signaling cascade. J Neurosci 16:909–918

Bielawski JP, Dunn KA, Sabehi G, Béàja O (2004) Darwinian adaptation of proteorhodopsin to different light intensities in the marine environment. Proc Natl Acad Sci USA 101:14824–14829

Del Punta K, Leinders-Zufall T, Rodriguez I, Jukam D, Wysocki CJ, Ogawa S, Zufall F, Mombaerts P (2002) Deficient pheromone responses in mice lacking a cluster of vomeronasal receptor genes. Nature 419:70–74

Del Punta K, Rothman A, Rodriguez I, Mombaerts P (2000) Sequence diversity and genomic organization of vomeronasal receptor genes in the mouse. Genome Res 10: 1958–1967

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

Felsenstein (1985) Confidence limits on phytogenies: an approach using the bootstrap. Evolution 39:783–791

Giorgi D, Friedman C, Trask BJ, Rouquier S (2000) Characterization of nonfunctional VIR-like pheromone receptor sequences in human. Genome Res 10:1979–1985

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–736

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phytogenies by maximum likelihood. Syst Biol 52(5):696–704

Halpern M (1987) The organization and function of the vomeronasal system. Annu Rev Neurosci 10:325–362

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174

Haydon DT, Bastos AD, Knowles NJ, Samuel AR (2001) Evidence for positive selection in foot-and-mouth disease virus capsid genes from field isolates. Genetics 157:7–15

Herrada G, Dulac C (1997) A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. Cell 90:763–737

Hughes AL, Hughes MK (1993) Adaptive evolution in the rat olfactory receptor gene family. J Mol Evol 36:249–54

Jia C, Halpern M (1996) Subclasses of vomeronasal receptor neurons: differential expression of G proteins (Gi alpha 2 and G(o alpha)) and segregated projections to the accessory olfactory bulb. Brain Res 719:117–128

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolisrm. Academic Press, New York, pp 21–132

Keverne EB (1999) The vomeronasal organ. Science 286:716–720

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 167:111–120

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17(12):1244–1245

Lane RP, Cutforth T, Axel R, Hood L, Trask BJ (2002) Sequence analysis of mouse vomeronasal receptor gene clusters reveals common promoter motifs and a history of recent expansion. Proc Natl Acad Sci USA 99:291–296

Lane RP, Young J, Newman T, Trask BJ (2004) Species specificity in rodent pheromone receptor repertoires. Genome Res 14:603–608

Matsunami H, Buck LB (1997) A multigene family encoding a diverse array of putative pheromone receptors in mammals. Cell 90:775–784

Metz EC, Robles-Sikisaka R, Vacquier VD (1998) Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. Proc Natl Acad Sci USA 95:10676–10681

Ngai J, Dowling MM, Buck L, Axel R, Chess A (1993) The family of genes encoding odorant receptors in the channel catfish. Cell 72:657–666

Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–936

O'Huigin C, Li WH (1992) The molecular clock ticks regularly in muroid rodents and hamsters. J Mol Evol 35:377–384

Pantages E, Dulac C (2000) A novel family of candidate pheromone receptors in mammals. Neuron 28:835–845

Rat Genome Sequencing Consortium(2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428:493–521

Rodriguez I, Del Punta K, Rothman A, Ishii T, Mombaerts P (2002) Multiple new and isolated families within the mouse superfamily of V1r vomeronasal receptors. Nat Neurosci 5:134–140

Rodriguez I, Greer CA, Mok MY, Mombaerts P (2000) A putative pheromone receptor gene expressed in human olfactory mucosa. Nat Genet 26:18–19

Rodriguez I, Mombaerts P (2002) Novel human vomeronasal receptor-like genes reveal species-specific families. Curr Biol 12:R409–R411

Roelofs WL, Rooney AP (2003) Molecular genetics and evolution of pheromone biosynthesis in Lepidoptera. Proc Natl Acad Sci USA 100:9179–9184

Ryba NJ, Tirindelli R (1997) A new multigene family of putative pheromone receptors. Neuron 19:371–379

Saito H, Mimmack ML, Keverne EB, Kishimoto J, Emson PC (1998) Isolation of mouse vomeronasal receptor genes and their co-localization with specific G-protein messenger RNAs. Brain Res Mol Brain Res 60:215–227

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

Sawyer S (1989) Statistical tests for detecting gene conversion. Mol Biol Evol 6:526–538

Shi P, Zhang J, Yang H, Zhang YP (2003) Adaptive diversification of bitter taste receptor genes in Mammalian evolution. Mol Biol Evol 20:805–814

Springer MS, Murphy WJ, Eizirik E, O'Brien SJ (2003) Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc Natl Acad Sci USA 100:1056–1061

Swanson WJ, Vacquier VD (2002) The rapid evolution of reproductive proteins. Nat Rev Genet 3:137–144

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

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Wysocki CJ, Lepri JJ (1991) Consequences of removing the vomeronasal organ. J Steroid Biochem Mol Biol 39:661–669

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555–556

Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. Trends Ecol Evol 15:496–503

Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908–917

Zhang J (2004) Frequent false detection of positive selection by the likelihood method with branch-site models. Mol Biol Evol 21:1332–1339

Zhang J, Webb DM (2003) Evolutionary deterioration of the vomeronasal pheromone transduction pathway in catarrhine primates. Proc Natl Acad Sci USA 100:8337–8341