Available online at www.sciencedirect.com

SCIENCE DIRECT

Genomics 86 (2005) 306 – 315

GENOMICS

www.elsevier.com/locate/ygeno

Composition and evolution of the V2r vomeronasal receptor gene repertoire in mice and rats

Hui Yang[a,b,1], Peng Shi[c,1], Ya-ping Zhang[a,d,*], Jianzhi Zhang[c,*]

[a]*Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China*
[b]*Graduate School of the Chinese Academy of Sciences, Beijing, China*
[c]*Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA*
[d]*Laboratory for Conservation and Utilization of Bio-resources, Yunnan University, Kunming, China*

Received 12 April 2005; accepted 10 May 2005
Available online 18 July 2005

Abstract

Pheromones are chemicals produced and detected by conspecifics to elicit social/sexual physiological and behavioral responses, and they are perceived primarily by the vomeronasal organ (VNO) in terrestrial vertebrates. Two large superfamilies of G protein-coupled receptors, V1rs and V2rs, have been identified as pheromone receptors in vomeronasal sensory neurons. Based on a computational analysis of the mouse and rat genome sequences, we report the first global draft of the V2r gene repertoire, composed of ~200 genes and pseudogenes. Rodent V2rs are subject to rapid gene births/deaths and accelerated amino acid substitutions, likely reflecting the species-specific nature of pheromones. Vertebrate V2rs appear to have originated twice prior to the emergence of the VNO in ancestral tetrapods, explaining seemingly inconsistent observations among different V2rs. The identification of the entire V2r repertoire opens the door to genomic-level studies of the structure, function, and evolution of this diverse group of sensory receptors.

Keywords: Evolution; Pheromone receptor; Rodents; Mouse; Rat; Vomeronasal; Gene family

Introduction

Pheromones are chemicals emitted and detected by individuals of the same species. They stimulate sexual and social changes in physiology and behavior, such as inducing estrus, terminating pregnancy, initiating copulatory behavior, and controlling intermale aggression [1]. Although some pheromones are detectable by the main olfactory system, most pheromones are probably sensed by the vomeronasal organ (VNO) in mammals [1]. VNO is encased in a bony capsule on the anterior nasal septum, and is anatomically and physiologically separated from the main olfactory system that detects thousands of odorants [1]. Two distinct superfamilies of seven-transmembrane G protein-coupled receptors, V1rs and V2rs, have been identified as vomeronasal pheromone receptors [2–5]. V1r genes have intronless coding regions. They are coexpressed with the G protein subunit $G\alpha_{i2}$ in sensory neurons whose cell bodies are located in the apical part of the vomeronasal epithelium [1,6]. In contrast, V2rs are characterized by the presence of a long, highly variable N-terminal domain. They are encoded by multiexon genes expressed in $G\alpha_O$-positive neurons whose cell bodies are located basally in the vomeronasal epithelium [1,6]. Neurons expressing V1r and V2r receptors project to the anterior and posterior accessory olfactory bulb, respectively, where they form multiple glomeruli in spatially conserved domains [1]. In addition to the distinct molecular structures and expression locations, V1r and V2r receptors also differ in a number of

* Corresponding authors. J. Zhang is to be contacted at Department of Ecology and Evolutionary Biology, University of Michigan, 1075 Natural Science Building, 830 North University Avenue, Ann Arbor, MI 48109, USA. Fax: +1 734 763 0544. Y. Zhang, Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China.
E-mail addresses: zhangyp1@263.net.cn (Y. Zhang), jianzhi@umich.edu (J. Zhang).
[1] These authors contributed equally to this work.

0888-7543/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2005.05.012

ways that suggest their different physiological functions. First, current experimental evidence shows that V1rs bind small volatile molecules, whereas V2rs bind peptides [7–11]. Second, gene knockout experiments suggest that V1rs are involved in gender discrimination, whereas V2rs control pheromone-induced male–male aggression [1,12–14]. Third, V2rs, but not V1rs, are coexpressed with M10 and M1 families of major histocompatibility complex (MHC) class Ib molecules in a selective fashion [14,15]. M10 molecules appear to function as escort molecules in the transport of V2rs to the cell membrane of vomeronasal sensory neurons [14].

With these differences, both V1rs and V2rs should be studied to gain a full understanding of the molecular mechanisms responsible for pheromone-based chemical communications. V1r genes are relatively easy to identify from genome sequences by computational methods because of their simple gene structures. To date, the complete V1r gene repertoire has been described in the human, chimpanzee, mouse, rat, dog, cow, and opossum [16–22], with the number of intact genes varying from a few in the human, chimpanzee, and dog to over 150 in the mouse. The evolution of the mammalian V1r repertoire is characterized by rapid gene turnover, lineage-specific phylogenetic clustering, accelerated nonsynonymous substitutions, and dramatic among-species variations of the repertoire size [10,18–24]. However, little is known about V2rs. Even in the model organisms of mouse and rat, only a few V2r genes have been described, although the total number of V2r genes is believed to be on the order of 100 [3–5]. This scarcity of knowledge is mainly due to the complex structure of V2r genes that makes their identification from genome sequences difficult. Here we combine several computational methods in an attempt to identify all V2r genes from the mouse and rat genome sequences. Subsequent analyses reveal both common features and unexpected differences between V1rs and V2rs in repertoire organization and evolution.

Results

Compositions of the mouse and rat V2r gene repertoires

To characterize the V2r gene repertoires in rodents, we searched the mouse and rat genome sequences using previously reported full-length V2r sequences as queries. Because V2r genes contain multiple exons (Fig. 1), we determined the exon/intron junctions by comparing the genomic sequences of newly identified V2r genes with cDNA sequences of known V2rs. Almost all exon/intron boundaries in our predicted genes were demarcated by standard donor/acceptor splice sites. As a result, 209 and 168 V2r genes were identified from the mouse and rat, respectively (Supplementary Tables 1–4), including 16 in mouse and 13 in rat that had been previously reported. The mouse genome sequence we used was essentially complete [25], while the rat genome sequence covered only 90% of genes [26]. Thus, the actual rat V2r repertoire may be slightly larger than reported here, and the number of V2r genes is probably similar between the two rodents. A total of 61 mouse and 57 rat V2r genes with intact ORFs were found (Table 1). Three V2rs (V2r52, V2r103, and V2r135) lacked the first ~200 amino acids in the N-terminal extracellular region, compared with canonical V2rs. These three short V2rs were tentatively regarded as functional, because one of them (V2r103) is expressed in VNO [3] and a homolog of metabotropic glutamate receptor (mGluR) with truncated N-terminus is known to function as an umami receptor [27]. The proportion of putatively functional members in the V2r repertoire is 29 and 34% for mouse and rat, respectively (Tables 2 and 3). These numbers are much lower than the corresponding numbers in the mouse and rat V1r repertoires (~50%) [17,19,20].

In mouse, V2r genes are distributed on 12 chromosomes, including chromosomes 1, 3, 5, 6, 7, 8, 10, 13, 14, 17, X, and Y (Table 2). The putatively functional genes, however, are located in 8 chromosomes, including chromosomes 3, 5, 6, 7, 10, 14, 17, and X. Rat V2r genes are located in 8 chromosomes, including 1, 2, 4, 5, 7, 12, 14, and 18, with chromosomes 5 and 14 having only pseudogenes (Table 3). In addition, there are 32 mouse genes that are still unmapped to chromosomes.

Phylogenetic relationships of rodent V2r genes

Mouse and rat V2rs are extremely variable at the protein sequence level. Fig. 1 illustrates the variability at each amino acid position along the protein sequence. Of the 950 amino acid positions in the alignment of the 118 intact V2rs of mouse and rat, 354 (37%) positions contain gaps and 298 positions (31%) have $\geq$10 different amino acids. The most variable positions, however, are almost exclusively located in the extracellular N-terminus, which is thought to be the ligand-binding domain [3–5]. To understand the evolutionary relationships among this diverse group of receptors, we reconstructed the phylogenetic tree of all putatively functional mouse and rat V2rs (Fig. 2). The V1r superfamily was previously classified into families based on the criterion that all members of a family must share at least 40% amino acid sequence identity [17]. Adopting the same criterion and using the phylogenetic information, we classified the intact V2r genes of the mouse and rat into families A, B, and C (Fig. 2; Table 1). The monophyly of each of the three families is supported by high bootstrap values (Fig. 2). Eighty-five percent of the mouse and 84% of the rat V2rs belong to family A. Family B contains only 6 mouse genes and 5 rat genes, whereas family C has 3 mouse genes and 4 rat genes. Most notably, family C includes the previously named V2r2 gene [5] that often exhibits features different from other V2rs (see below). The average protein sequence identity between family A and family B is 38%, whereas

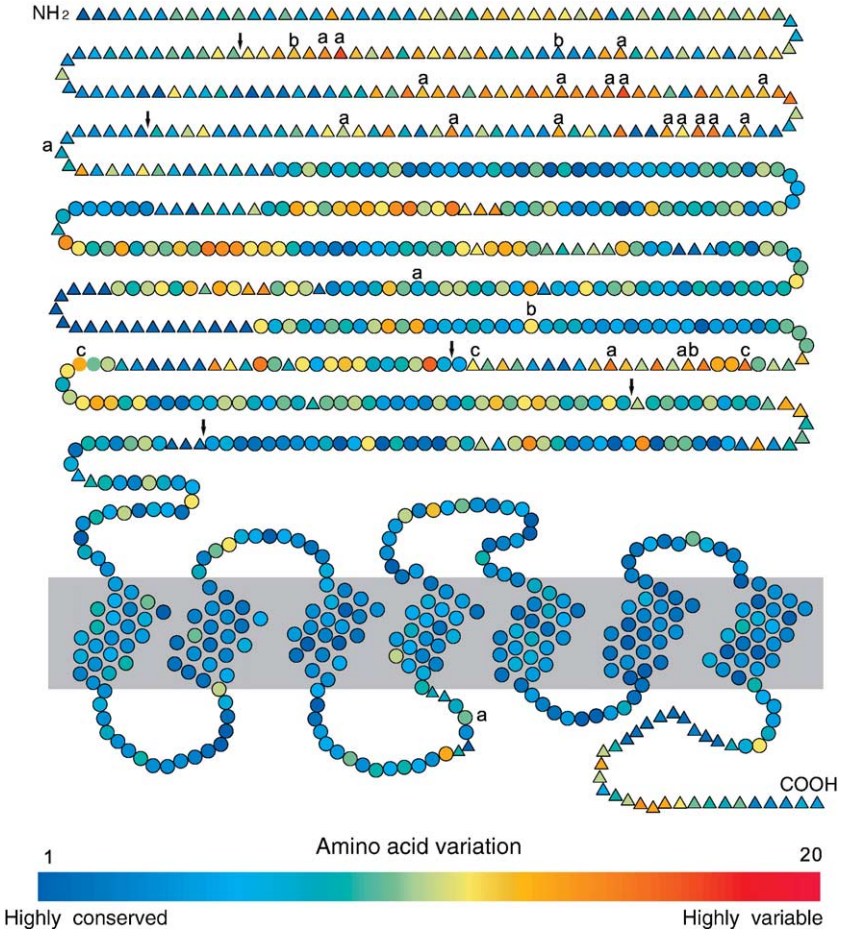

Fig. 1. Sequence variation of V2rs. Color shows the number of different amino acids (from 1 to 20) observed at each position of the alignment of all 118 intact V2rs from the mouse and rat. A circle indicates that no gap is observed at the position, whereas a triangle indicates the observation of a gap in the alignment. Arrows show positions of introns. The pink shaded area represents the cell membrane. Positively selected sites (with Bayes Empirical Bayes posterior probability >0.99) are marked with a, b, or c, indicating the family identity of V2rs (A, B, or C) in which positive selection was detected. The sequence alignment is presented in Supplementary Fig. 1.

that between family C and families A and B is 27%. Because family A contains a total of 52 mouse and 48 rat genes, one may further divide the family into 9 clades (I–IX) based on the phylogeny. Although this division is somewhat arbitrary and is only used to help describe the evolution of the V2r genes, all nine clades have strong bootstrap support (Fig. 2). Except for clades II, VII, and VIII, all clades contain both mouse and rat genes, indicating that the 9 clades arose before the mouse–rat separation. Thus, the absence of mouse genes in clades II and VII and the absence of rat genes in clade VIII likely resulted from lineage-specific gene losses. Notably, in both mouse and rat, V2r genes that cluster in the phylogenetic tree also tend to cluster in chromosomal locations (see Supplementary Tables 1 and 2), revealing tandem gene duplication as the primary source of new V2r genes in evolution.

Table 1
Summary statistics of mouse and rat V2r genes

Family	Number of intact genes		Protein sequence identity (%)	
	Mouse	Rat	Mouse	Rat
Family A	52	48	55–99	54–96
Family B	6	5	72–91	78–90
Family C	3	4	81–86	78–86
Total	61 (148)	57 (111)		

Numbers of pseudogenes are given in parentheses.

The phylogenetic tree also shows that the mouse and rat V2r genes cluster largely by species, forming many species-specific clades. In fact, only four pairs of one-to-one orthologs (V2r49 and V2r133; V2r55 and V2r101; V2r21 and V2r108; V2r63 and V2r148) are found in the entire tree. The species-specific clades of V2r genes are likely due to recent gene duplications and/or gene losses that took place after the mouse–rat separation. Alternatively, they may have resulted from frequent gene conversions that have homogenized gene sequences within species. We used Sawyer's method [28] to examine gene conversion among paralogous genes within species. Among 61 mouse V2r genes, 14 pairs were found to exhibit gene conversion at the 5% significance level (after the Bonferroni correction for multiple tests). The signals of gene conversion, however, were mainly from 5

Table 2
Chromosomal distributions of mouse V2r loci

Cytogenetic position	No. of intact genes	No. of pseudogenes	Percentage of intact genes
1qH6	0	1	0
3qE1	3	5	38
5qF	7	5	58
5qG2	0	1	0
5qG3	0	3	0
5_random[a]	1	0	100
6qF2	6	3	67
7qA1	13	12	52
7qB1	3	7	30
7qB4	0	1	0
7qD2	5	9	36
7qD3	1	2	33
7_random[b]	0	5	0
8qA1.1	0	2	0
10qC1	4	1	80
10qD3	2	2	50
13qA3.3	0	1	0
13qB3	0	2	0
14qC1	2	2	50
17qA3.1	2	2	50
17qA3.2	7	20	26
17qA3.3	0	2	0
17qC	0	1	0
17qD	1	0	100
XqE3	1	0	100
YqA1	0	8	0
YqA2	0	6	0
YqB	0	3	0
YqC1	0	1	0
YqC2	0	1	0
YqC3	0	3	0
YqD	0	6	0
YqE	0	2	0
Un_random[c]	3	29	9
Sum	61	148	29

[a] Exact location on chromosome 5 is unknown.
[b] Exact location on chromosome 7 is unknown.
[c] Location in the genome is unknown.

genes, because gene conversion was no longer found when the 5 genes were removed. Similarly, among the 57 rat V2rs, 19 pairs were detected to have undergone gene conversion. Gene conversion was no longer found after we removed 8 genes. Furthermore, significance in Sawyer's test could be due to nonrandom distribution of nucleotide substitutions along the sequences, rather than gene conversion. Thus, gene conversion was a minor factor in the evolution of rodent V2r genes and most species-specific gene clades were results of recent gene duplications and/or losses.

To date the gene duplication events in the mouse and rat V2r superfamilies, we first removed the 5 mouse and 8 rat genes that had gene conversions. With the assumption that the mouse and rat diverged 18 million years (MY) ago [26], we estimated that the synonymous substitution rate is $0.19/(18 \times 10^6 \times 2) = 5.3 \times 10^{-9}$ per site per year in rodents, because the genome-wide average number of synonymous nucleotide substitutions per site is 0.19 between the mouse and rat [26]. We limited our molecular dating to the duplication events within families A, B, and C, as these events were relatively recent and the use of the above rodent synonymous rate for calibration is justifiable. The synonymous substitution rate is likely to be substantially higher in rodents than in other vertebrates [26] and use of the rodent rate outside rodents may seriously underestimate divergence times. We found that the first duplication event within family A occurred ~86 MY ago and all 9 clades in family A were present by ~54 MY ago. Because the sister lineages of Rodentia and Lagomorpha diverged ~85 MY ago [29], it is likely that the expansion of family A occurred within rodents. We found that many duplication events within mouse and rat V2r families postdated the mouse–rat separation. Interestingly, about 60% of the gene duplications within the three families of mouse V2rs occurred after the mouse–rat split, compared with 46% in rat (Fig. 3). About 45% of the gene duplications took place in the last 10 MY in mouse, compared with 33% in rat. These dating results are consistent with our phylogenetic analysis (Fig. 2) and suggest that the rate of successful gene duplication in V2r families is higher in the mouse lineage than in the rat lineage after their separation.

Possible independent origins of family C and families A and B

Our phylogenetic analysis showed that V2rs of family C are very divergent from those of families A and B. To gain a better understanding of the timing of this divergence, we analyzed the mouse and rat intact V2r genes with previously reported V2rs from the goldfish, fugu, and African clawed frog. In addition, we identified new V2r genes from the genome sequences of the zebrafish, puffer fish, fugu, and

Table 3
Chromosomal distributions of rat V2r loci

Cytogenetic position	No. of intact genes	No. of pseudogenes	Percentage of intact genes
1p13	4	14	22
1p12	0	1	0
1q12	14	29	33
1q21	5	13	28
1q22	0	4	0
1q31	0	1	0
1q32	5	3	63
2q31	4	4	50
4q42	5	1	83
5q31	0	1	0
7q11	3	2	60
7q12	1	1	50
7q13	2	1	67
12p12	2	9	18
12q11	3	4	43
12q12	0	9	0
14p22	8	12	40
14q11	0	1	0
18p13	1	1	50
Sum	57	111	34

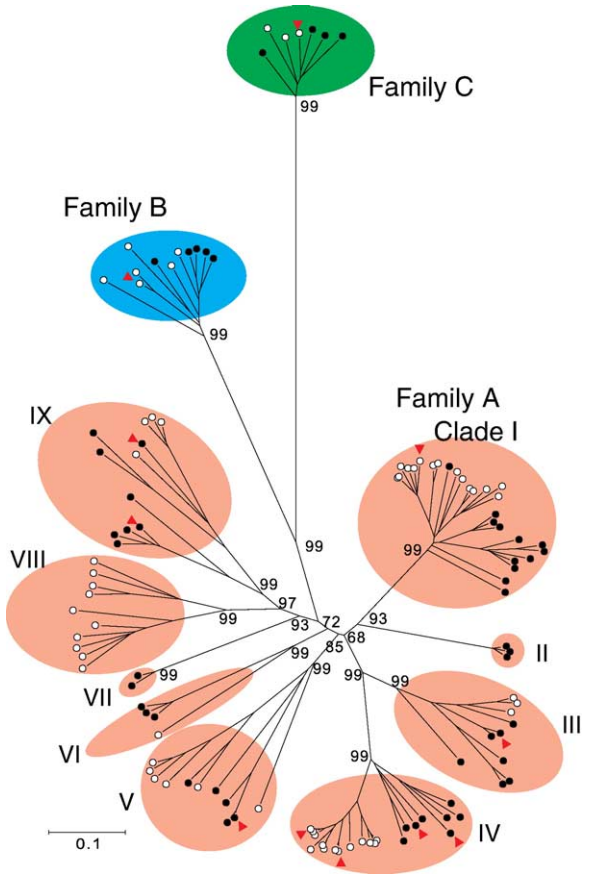

Fig. 2. Phylogenetic relationships of mouse and rat intact V2r genes. The tree is reconstructed by the neighbor-joining method with protein Poisson distances. Bootstrap percentages for major nodes are shown. The same tree with bootstrap percentages at all nodes is presented in Supplementary Fig. 2. The mouse and rat V2r genes are represented by open and closed circles, respectively. Red arrows indicate those genes known to be expressed in the VNO. The three families are shown with different colors. See text for definitions of families and clades. The scale bar shows 0.1 amino acid substitution per site.

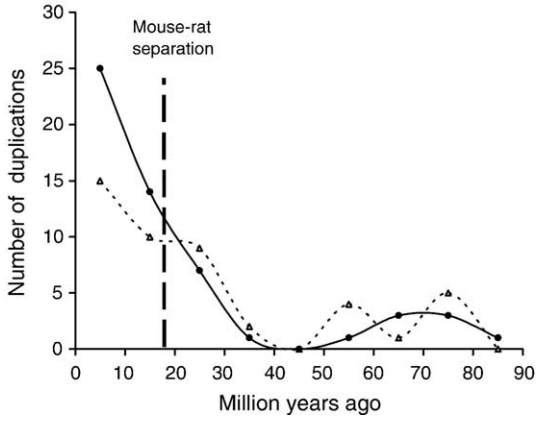

Fig. 3. Number of duplications within each 10 MY interval in the evolution of mouse and rat V2r superfamilies. Duplication events within families are considered, after the removal of genes involved in gene conversion. The solid line is for mouse V2rs, whereas the dotted line is for rat V2rs. Mouse and rat diverged ~18 MY ago [26], as indicated by the vertical line.

western clawed frog and found unpublished carp sequences in GenBank. V2rs belong to family 3/C G-protein-coupled receptors, which include Ca^{2+}-sensing receptors (CasRs) and taste receptors (TAS1Rs), among other receptors [30]. Therefore, we also included multiple sequences of these receptors in our analysis. The phylogenetic tree shows several interesting branching patterns (Fig. 4). First, rodent V2rs of families A and B cluster together with a 97% bootstrap value, in exclusion of frog and fish V2rs, suggesting that the divergence of family A and family B postdated the separation of mammals from amphibians. Second, there are multiple V2r genes from frogs and fishes and they can be classified into two phylogenetic groups, with one group clustering with mouse and rat families A and B, and the other clustering with family C. This pattern indicates that the divergence of family C and families A and B predated the separation of tetrapods and teleost fish. Most

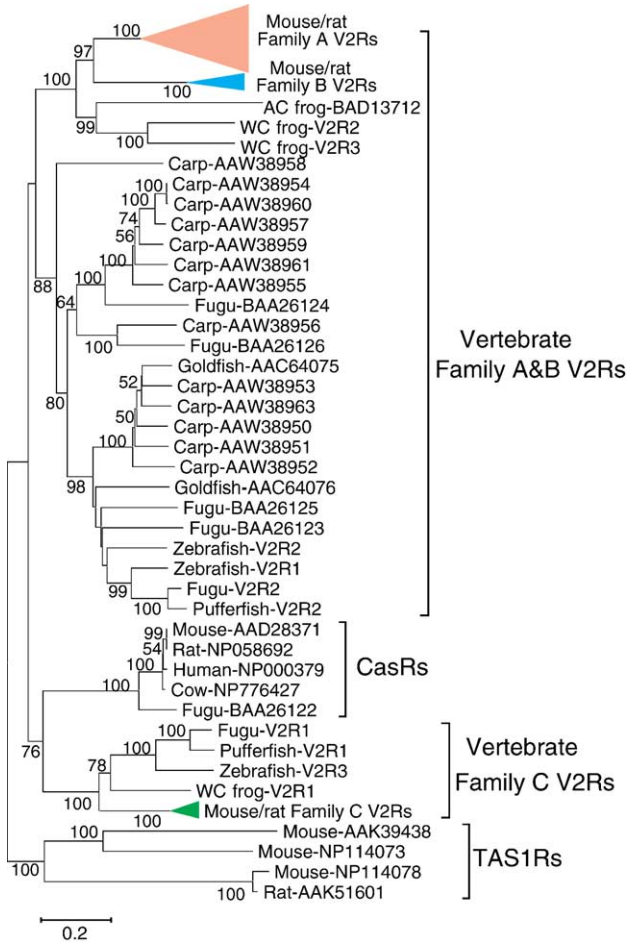

Fig. 4. Origin and evolution of vertebrate V2r genes. The tree is reconstructed by the neighbor-joining method with protein Poisson distances. Bootstrap percentages higher than 50 are shown. AC frog, African clawed frog. WC frog, western clawed frog. Taste receptors (TAS1R) are used as outgroups [30]. Families of mouse and rat V2r genes are collapsed for better illustration. GenBank accession numbers are shown for those genes already in GenBank, whereas gene names are provided for those identified in this study. The scale bar shows 0.2 amino acid substitution per site.

surprisingly, vertebrate V2rs of family C are evolutionarily closer to CasRs than to V2rs of families A and B. Note that this relationship has a moderate bootstrap support (76%), suggesting the intriguing scenario that family C and families A and B had independent origins.

Selective pressures during the divergence of rodent V2rs

Positive selection was suggested to have driven the divergence of V2r genes in an analysis of a small number of mouse and rat genes [10]. It is of interest to examine the role of positive selection on the entire V2r repertoire. We calculated the number of synonymous (d_S) and nonsynonymous (d_N) substitutions per site among paralogous genes within each of the three mouse V2r families. In no comparisons did we find $d_N > d_S$. The mean d_N for the N-terminal extracellular region is significantly greater than the mean d_N for the rest of the protein ($P < 0.001$). But there is no difference in mean d_S between these two regions. The same is found for the rat V2r genes. It is interesting to note that the mean d_N/d_S ratio in the non-N-terminal region of rodent V2rs (0.28–0.29) is similar to the d_N/d_S ratio in an average rodent gene (0.23) [31], while the mean d_N/d_S ratio in the N-terminal region is 2–3 times higher (Supplementary Fig. 3). This high rate suggests relatively weak purifying selection and/or positive selection acting in the N-terminal region. Using the likelihood method, we tested the presence of positive selection in rodent V2rs by examining family B, family C, and each of the nine clades of family A separately. As shown in Supplementary Table 5, positive selection was detected in family B, familiy C, and clades I, III, IV, VI, and IX of family A. The 27 amino acid positions detected to be under positive selection were indicated in Fig. 1. As expected, most of these sites show high variation and all sites except one are located in the N-terminal extracellular region.

Discussion

In this study, we provided the first global draft of the mammalian V2r repertoire by searching the mouse and rat genome sequences. A total of 209 and 168 V2r genes were identified, including 61 and 57 putatively functional genes, in mouse and rat, respectively. These numbers are consistent with earlier estimates based on hybridization experiments [3,4,32]. Mouse has 187 intact V1r genes [21] and 165 pseudogenes [20], whereas rat has 106 intact V1r genes and 110 pseudogenes [20]. Thus, in both mouse and rat, the V1r repertoire appears much larger than the V2r repertoire, particularly with regard to intact genes. One possible explanation of this difference is underdetection of V2rs due to the difficulty in their computational identification. However, because pseudogenes are more difficult to detect than intact genes, the methodological imperfection would lead to overestimation of the proportion of intact genes among all genes detected. But, we observed a lower proportion of intact genes in the V2r repertoire than in the V1r repertoire, arguing against the above explanation. In fact, the number of pseudogenes is similar in V1r and V2r gene superfamilies for both mouse and rat. We thus believe that the observed repertoire size difference between V1rs and V2rs is real.

Based on the same family definition, V1rs of mouse and rat can be divided into 15 families [18,21,22], whereas V2rs can be classified into only three families. Thus, at the family level, V1rs are much more diverse than V2rs. However, all 15 V1r families appear to have diversified after the separation of mammals and amphibians ([22]; W. Grus and J. Zhang, unpublished), whereas the divergence of V2r family C and families A and B predated the separation of tetrapods and teleost fish. Thus, at the superfamily level, mammalian V2rs may be considered more diverse than V1rs.

The VNO is believed to have originated in an ancestral tetrapod and it does not exist in fish [33]. But V2r genes are known to exist in goldfish [34]. In the present study, we also identified V2r genes in puffer fish and zebrafish. Fish V2r genes are expressed in a subset of olfactory receptor cells and some fish V2rs are activated primarily by amino acids, which are fish odorants [35]. It is currently unknown whether a primordial vomeronasal sensory system exists in fish. We found that both family A and B-like and family C-like V2r genes exist in fish and amphibians, strongly suggesting the presence of both types of V2r genes in the common ancestor of tetrapods and teleost fish. More surprisingly, our phylogenetic analysis suggests the possibility that the two types of V2rs had independent evolutionary origins. This explains previous findings of several differences between the two types of V2rs. First, studies of V2rs of families A and B demonstrated that only one gene is expressed per vomeronasal sensory neuron [3,4,36]. A subsequent study, however, showed that family C V2r coexpresses with other V2r genes [36]. Second, V2rs of families A and B, but not family C, are coexpressed with M10 and M1 families of MHC class Ib molecules [14,15,37]. M10 molecules appear to function as escort molecules in transport of V2rs of families A and B, but not family C, to the cell membrane of vomeronasal sensory neurons [14,37]. These different expression patterns, together with the possibility of independent evolutionary origins, suggest that family C V2rs may be functionally distinct from families A and B.

The evolution of the rodent V1r repertoire is characterized by rapid gene turnover, lineage-specific phylogenetic clustering, and accelerated nonsynonymous substitutions [10,18–22]. All these features are also observed for the V2r repertoire. For instance, only four pairs of one-to-one V2r orthologs are found between the mouse and the rat and almost all V2r genes form species-specific gene clades. These patterns likely resulted from rapid gene sorting, a process of differential retention, and amplification of ancestral genes [38] via gene births and deaths [39,40].

Our molecular dating also shows that most duplication events within mouse and rat V2r families occurred after the separation of the two species. Interestingly, species-specific gene clades of mouse and rat have also been described for bitter taste receptor genes and olfactory receptor genes [41–43]. These species-specific sensory receptors possibly detect chemical stimuli uniquely encountered by each species [18,21,42]. The rate of successful gene duplications in the mouse and rat V2r superfamily is surprisingly high, particularly in the past 10 MY. This high rate of successful gene duplication might be due to a high rate of mutation that generates the duplication and/or a high probability of fixation and retention of the duplicated genes. Lane et al. suggested that the mutations that generate V1r duplicate genes in mice were mediated by L1 repetitive elements, based on the observation of high L1 density in the genomic regions harboring V1r genes [44]. We found that the L1 density of the genomic regions containing the majority of V2r genes is 43% in mouse and 37% in rat, much higher than the genomic average density of 20% in mouse and 23% in rat [26]. Although V1rs and V2rs are not distributed in the same chromosomal regions, the above analysis suggests a similar role of L1 elements in V1r and V2r duplications.

Previous maximum-likelihood analysis suggested that positive selection plays an important role in the divergence of rodent V1r genes [10,21,24]. However, application of a more conservative method did not find convincing evidence for positive selection [18,23]. In V2rs, the average d_N/d_S ratio in the N-terminal extracellular region is significantly greater than that in the rest of the protein. However, we did not find higher d_N than d_S in any pairwise comparison of paralogous V2rs when d_N and d_S were averaged for all sites of the N-terminal region. Our subsequent likelihood analysis identified 27 amino acid sites that show signals of positive selection (Supplementary Table 5; Fig. 1). As expected, almost all of these sites are located in the N-terminal region, which is thought to be the ligand-binding domain [3–5]. A similar result was previously obtained in an analysis of 9 V2r genes [10]. We caution that the likelihood method is known to make false detection of positive selection under certain conditions [45,46]. The positively selected sites reported here and elsewhere should be confirmed by other statistical methods and/or experiments. These limitations notwithstanding, if the identified amino acid sites in the N-terminal region are indeed subject to positive selection, they may determine binding specificity and provide useful information on the diversity of pheromones within and between species. An interesting observation from our analysis is that family A, B, and C V2rs have different sets of positively selected sites, suggesting the possibility that the three families bind structurally distinct classes of ligands.

V1rs and V2rs show many differences in their expression and function. The identification of the entire V2r repertoire from the mouse and rat opens the possibility for genomic level studies of the structure, function, and evolution of V2rs, a necessary step toward the full understanding of the molecular mechanisms underlying pheromone communications.

Materials and methods

Identification of V2r genes

Sequences of previously reported mouse and rat V2r genes were retrieved from the GenBank or literature, including VR1 to VR16 (Accession Numbers AF011411-AF011426) [4], Go-VN1 to Go-VN7, Go-VN13C (AF016178-AF016185) [3], V2r2 (NM_019918) [5], Mm_V2r1a to Mm_V2r1d, and Rn_V2r1a to Rn_V2r1e [10]. Additional V2r genes were obtained by searching the mouse *Mus musculus* (May 2004 assembly) and rat *Rattus norvegicus* (November 2004 assembly) genome sequences at UCSC (http://www.genome.cse.ucsc.edu/) and Ensembl (http://www.ensembl.org/), respectively. To ensure the accurate prediction of V2r genes, we used a computational strategy involving multiple steps. First, those previously reported sequences with intact open-reading frames (ORFs), including the N-terminal extracellular domain and the seven transmembrane domains, were used as query sequences. Second, these query sequences were used to identify the genomic locations of homologous genes in a genome by the BLAST program [47]. Third, the genomic DNA sequences of the homologous genes and the known V2r cDNA sequences were used to conduct cDNA-to-genomic sequence alignment on Spidey (http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/), which provided the exon/intron structures and full-length protein sequences of the homologous genes. Spidey is known to perform well even when confronted with noise from alternative splicing, polymorphisms, sequencing errors, and evolutionary divergence [48]. Fourth, the obtained putative protein sequences were examined by the TMHMM method [49] for the presence of seven transmembrane domains. Finally, the identified putative V2r genes were BLASTed against the nr database of GenBank to ensure that the best hit was a V2r. This step was necessary because some other receptors (e.g., Ca^{2+}-sensing receptors; TAS1R taste receptors) are known to be homologous to V2rs. It should be noted that some previously described V2r sequences did not match exactly with the genome sequences. In such cases, gene sequences derived from the genome sequences were used, with the previously used gene names also being provided (Supplementary Tables 1 and 2). These sequence differences might have resulted from errors in the genome sequences, errors in previously reported sequences, or polymorphisms due to the use of different mouse or rat strains by different groups. Sensory genes commonly have high levels of intraspecific polymorphisms [19,43]. A V2r sequence was regarded as a pseudogene if its disrupted open-reading frame was longer than 200 nucleotides, which could usually code for two transmembrane domains and a connecting loop. This strin-

gent criterion led to a low rate of false detection, although some true V2r pseudogenes might have been missed.

To avoid confusion and to establish a uniform nomenclature system for V2rs, we adopted the official V2r nomenclature for them. For those previously reported V2r genes, the original names were used. Otherwise, V2r# was used, where # can be from 1 to 100 for a mouse and from 101 to 200 for a rat gene. In Supplementary Tables 1 and 2, we list all V2r genes of the mouse and rat, respectively, with the corresponding gene names provided. In Supplementary Tables 3 and 4, we list all V2r pseudogenes of the mouse and rat, respectively. The physical locations of the intact genes and pseudogenes in the mouse and rat genomes are provided in Supplementary Tables 1–4 as well. In addition, the information of the newly obtained sequences is also available in Mouse Genome Informatics (MGI) database (www.informatics.jax.org).

The presence of V2r genes has also been reported in the African clawed frog *Xenopus laevis*, goldfish *Carassius auratus*, and fugu *Takifugu rubripes* [34,50,51]. These fish and frog sequences were used along with our identified mouse and rat V2rs as queries to search the genome sequences of the western clawed frog *Xenopus tropicalis*, zebrafish *Danio rerio*, puffer fish *Tetraodon nigroviridis* and fugu through UCSC (http://genome.cse.ucsc.edu/). Although the puffer fish genome sequence had an 8.3-fold coverage, the frog, zebrafish, and fugu genome sequences had low coverages. We thus did not intend to recover the entire V2r superfamily in these four species. Instead, the purpose was to obtain a few V2r sequences from non-mammalian vertebrates to aid the phylogenetic analysis of mouse and rat V2rs, as the mammalian V2rs appear to have multiple ancient origins (see Results).

Evolutionary analyses

Deduced amino acid sequences of V2rs were aligned by CLUSTAL_X [52], with manual adjustments. Supplementary Fig. 1 shows the alignment of the 118 mouse and rat intact V2rs. A phylogenetic tree of these V2rs was reconstructed using the neighbor-joining method [53] with protein Poisson distances [54]. Gap sites in the alignment were not used in the phylogenetic reconstruction (the complete-deletion option). The reliability of the estimated tree was evaluated by the bootstrap method [55] with 1000 pseudo-replications. MEGA2 [56] was used for the phylogenetic analysis. Sawyer's method [28] as implemented in the computer program GENECONV 1.81 was used to examine gene conversion among paralogous genes in mouse. Exons within genes were concatenated for this analysis and the default parameters were used. Because our gene trees were reconstructed using exon sequences only, gene conversion in introns would not interfere with our phylogenetic analysis and therefore was not considered here. Among 1830 possible pairs of the 61 mouse V2r genes, 14 pairs were found to exhibit gene conversion at the 5% significance level (after the Bonferroni correction for multiple tests). If gene conversion occurs more frequently among closely related sequences, the above analysis will underestimate gene conversion due to the global Bonferroni correction. We thus conducted an analysis for each phylogenetic clade in Fig. 2. However, this analysis only identified 9 pairs of genes that exhibit gene conversion, suggesting that the global analysis was sufficiently powerful. Similarly, among the 1596 possible pairs of the 57 rat V2rs, 19 pairs were detected to have undergone gene conversion. A clade-by-clade analysis showed only 17 pairs of genes with signals of gene conversion. We thus report the global analysis results only. The distribution of L1 repetitive elements in genomic sequences was determined by the RepeatMasker program (http://www.repeatmasker.org/), following the same criterion used in [44]. Specifically, we estimated the L1 density in intergenic regions for tandem-linked intact V2r genes and in 10,000 nucleotides upstream and downstream of the coding regions for non-tandem-linked intact V2r genes. To examine the pattern of nucleotide substitutions, the number of synonymous substitutions per synonymous site (d_S) and the number of nonsynonymous substitutions per nonsynonymous site (d_N) between two homologous sequences were estimated by the modified Nei-Gojobori method [57]. In addition, the maximum-likelihood method [58] as implemented in PAML [59] was used to test positive selection at individual codons. To date the gene duplication events within V2r families, we computed d_S between a pair of paralogous genes or two groups of paralogous genes. A mean synonymous substitution rate of 5.3×10^{-9} per site per year, derived from the mouse-rat genomic comparison [26], was used for calibration.

Acknowledgments

We thank Soochin Cho, Wendy Grus, and David Webb for valuable comments and Deng Pan for computational assistance with Fig. 1. This work was supported in part by a National Institutes of Health grant (GM67030) to J.Z. and grants from National Natural Science Foundation of China (30021004 and 30430110), Natural Science Foundation of Yunnan, and Chinese Academy of Sciences to Y.P.Z.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2005.05.012.

References

[1] C. Dulac, A.T. Torello, Molecular detection of pheromone signals in mammals: from genes to behaviour, Nat. Rev. Neurosci. 4 (2003) 551–562.

[2] C. Dulac, R. Axel, A novel family of genes encoding putative pheromone receptors in mammals, Cell 83 (1995) 195–206.
[3] G. Herrada, C. Dulac, A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution, Cell 90 (1997) 763–773.
[4] H. Matsunami, L.B. Buck, A multigene family encoding a diverse array of putative pheromone receptors in mammals, Cell 90 (1997) 775–784.
[5] N.J. Ryba, R. Tirindelli, A new multigene family of putative pheromone receptors, Neuron 19 (1997) 371–379.
[6] P. Mombaerts, Genes and ligands for odorant, vomeronasal and taste receptors, Nat. Rev. Neurosci. 5 (2004) 263–278.
[7] T. Leinders-Zufall, A.P. Lane, A.C. Puche, W. Ma, M.V. Novotny, M.T. Shipley, et al., Ultrasensitive pheromone detection by mammalian vomeronasal neurons, Nature 405 (2000) 792–796.
[8] K. Del Punta, T. Leinders-Zufall, I. Rodriguez, D. Jukam, C.J. Wysocki, S. Ogawa, et al., Deficient pheromone responses in mice lacking a cluster of vomeronasal receptor genes, Nature 419 (2002) 70–74.
[9] C. Boschat, C. Pelofi, O. Randin, D. Roppolo, C. Luscher, M.C. Broillet, et al., Pheromone detection mediated by a V1r vomeronasal receptor, Nat. Neurosci. 5 (2002) 1261–1262.
[10] R.D. Emes, S.A. Beatson, C.P. Ponting, L. Goodstadt, Evolution and comparative genomics of odorant- and pheromone-associated genes in rodents, Genome Res. 14 (2004) 591–602.
[11] T. Leinders-Zufall, P. Brennan, P. Widmayer, P. Chandramani, A. Maul-Pavicic, M. Jager, et al., MHC class I peptides as chemosensory signals in the vomeronasal organ, Science 306 (2004) 1033–1037.
[12] L. Stowers, T.E. Holy, M. Meister, C. Dulac, G. Koentges, Loss of sex discrimination and male-male aggression in mice deficient for TRP2, Science 295 (2002) 1493–1500.
[13] B.G. Leypold, C.R. Yu, T. Leinders-Zufall, M.M. Kim, F. Zufall, R. Axel, Altered sexual and social behaviors in trp2 mutant mice, Proc. Natl. Acad. Sci. USA 99 (2002) 6376–6381.
[14] J. Loconto, F. Papes, E. Chang, L. Stowers, E.P. Jones, T. Takada, et al., Functional expression of murine V2R pheromone receptors involves selective association with the M10 and M1 families of MHC class Ib molecules, Cell 112 (2003) 607–618.
[15] T. Ishii, J. Hirota, P. Mombaerts, Combinatorial coexpression of neural and immune multigene families in mouse vomeronasal sensory neurons, Curr. Biol. 13 (2003) 394–400.
[16] I. Rodriguez, P. Mombaerts, Novel human vomeronasal receptor-like genes reveal species-specific families, Curr. Biol. 12 (2002) R409–R411.
[17] I. Rodriguez, K. Del Punta, A. Rothman, T. Ishii, P. Mombaerts, Multiple new and isolated families within the mouse superfamily of V1r vomeronasal receptors, Nat. Neurosci. 5 (2002) 134–140.
[18] W.E. Grus, J. Zhang, Rapid turnover and species-specificity of vomeronasal pheromone receptor genes in mice and rats, Gene 340 (2004) 303–312.
[19] X. Zhang, I. Rodriguez, P. Mombaerts, S. Firestein, Odorant and vomeronasal receptor genes in two mouse genome assemblies, Genomics 83 (2004) 802–811.
[20] J.M. Young, M. Kambere, B.J. Trask, R.P. Lane, Divergent V1R repertoires in five species: amplification in rodents, decimation in primates, and a surprisingly small repertoire in dogs, Genome Res. 15 (2005) 231–240.
[21] P. Shi, J.P. Bielawski, H. Yang, Y.P. Zhang, Adaptive diversification of vomeronasal receptor 1 genes in rodents, J. Mol. Evol. (in press).
[22] W.E. Grus, P. Shi, Y.P. Zhang, J. Zhang, Dramatic variation of the vomeronasal pheromone receptor gene repertoire among five orders of placental and marsupial mammals, Proc. Natl. Acad. Sci. USA 102 (2005) 5767–5772.
[23] J. Zhang, D.M. Webb, Evolutionary deterioration of the vomeronasal pheromone transduction pathway in catarrhine primates, Proc. Natl. Acad. Sci. USA 100 (2003) 8337–8341.
[24] N.I. Mundy, S. Cook, Positive selection during the diversification of class I vomeronasal receptor-like (V1RL) genes, putative pheromone receptor genes, in human and primate evolution, Mol. Biol. Evol. 20 (2003) 1805–1810.
[25] R.H. Waterston, K. Lindblad-Toh, E. Birney, J. Rogers, J.F. Abril, P. Agarwal, et al., Initial sequencing and comparative analysis of the mouse genome, Nature 420 (2002) 520–562.
[26] R.A. Gibbs, G.M. Weinstock, M.L. Metzker, D.M. Muzny, E.J. Sodergren, S. Scherer, et al., Genome sequence of the Brown Norway rat yields insights into mammalian evolution, Nature 428 (2004) 493–521.
[27] N. Chaudhari, A.M. Landin, S.D. Roper, A metabotropic glutamate receptor variant functions as a taste receptor, Nat. Neurosci. 3 (2000) 113–119.
[28] S. Sawyer, Statistical tests for detecting gene conversion, Mol. Biol. Evol. 6 (1989) 526–538.
[29] W.J. Murphy, P.A. Pevzner, S.J. O'Brien, Mammalian phylogenomics comes of age, Trends Genet. 20 (2004) 631–639.
[30] J.P. Pin, T. Galvez, L. Prezeau, Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors, Pharmacol. Ther. 98 (2003) 325–354.
[31] J. Zhang, Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes, J. Mol. Evol. 50 (2000) 56–68.
[32] H. Kouros-Mehr, S. Pintchovski, J. Melnyk, Y.J. Chen, C. Friedman, B. Trask, et al., Identification of non-functional human VNO receptor genes provides evidence for vestigiality of the human VNO, Chem. Senses 26 (2001) 1167–1174.
[33] H.L. Eisthen, Phylogeny of the vomeronasal system and of receptor cell types in the olfactory and vomeronasal epithelia of vertebrates, Microsc. Res. Tech. 23 (1992) 1–21.
[34] Y. Cao, B.C. Oh, L. Stryer, Cloning and localization of two multigene receptor families in goldfish olfactory epithelium, Proc. Natl. Acad. Sci. USA 95 (1998) 11987–11992.
[35] D.J. Speca, D.M. Lin, P.W. Sorensen, E.Y. Isacoff, J. Ngai, A.H. Dittman, Functional identification of a goldfish odorant receptor, Neuron 23 (1999) 487–498.
[36] S. Martini, L. Silvotti, A. Shirazi, N.J. Ryba, R. Tirindelli, Co-expression of putative pheromone receptors in the sensory neurons of the vomeronasal organ, J. Neurosci. 21 (2001) 843–848.
[37] L. Silvotti, G. Giannini, R. Tirindelli, The vomeronasal receptor V2R2 does not require escort molecules for expression in heterologous systems, Chem. Senses 30 (2005) 1–8.
[38] J. Zhang, K.D. Dyer, H.F. Rosenberg, Evolution of the rodent eosinophil-associated RNase gene family by rapid gene sorting and positive selection, Proc. Natl. Acad. Sci. USA 97 (2000) 4701–4706.
[39] A.L. Hughes, M. Nei, Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals, Mol. Biol. Evol. 6 (1989) 559–579.
[40] M. Nei, A.L. Hughes, in: K. Tsuji, M. Aizawa, T. Sasazuki (Eds.), Proceedings of the 11th Histocompatibility Workshop and Conference, Oxford Univ. Press, Oxford, 1992, pp. 27–38.
[41] X. Zhang, S. Firestein, The olfactory receptor gene superfamily of the mouse, Nat. Neurosci. 5 (2002) 124–133.
[42] P. Shi, J. Zhang, H. Yang, Y.P. Zhang, Adaptive diversification of bitter taste receptor genes in mammalian evolution, Mol. Biol. Evol. 20 (2003) 805–814.
[43] X. Wang, S.D. Thomas, J. Zhang, Relaxation of selective constraint and loss of function in the evolution of human bitter taste receptor genes, Hum. Mol. Genet. 13 (2004) 2671–2678.
[44] R.P. Lane, T. Cutforth, R. Axel, L. Hood, B.J. Trask, Sequence analysis of mouse vomeronasal receptor gene clusters reveals common promoter motifs and a history of recent expansion, Proc. Natl. Acad. Sci. USA 99 (2002) 291–296.
[45] Y. Suzuki, M. Nei, Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites, Mol. Biol. Evol. 19 (2002) 1865–1869.
[46] J. Zhang, Frequent false detection of positive selection by the

likelihood method with branch-site models, Mol. Biol. Evol. 21 (2004) 1332–1339.
[47] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[48] S.J. Wheelan, D.M. Church, J.M. Ostell, Spidey: a tool for mRNA-to-genomic alignments, Genome Res. 11 (2001) 1952–1957.
[49] E.L. Sonnhammer, G. von Heijne, A. Krogh, A hidden Markov model for predicting transmembrane helices in protein sequences, Proc. Int. Conf. Intell. Syst. Mol. Biol. 6 (1998) 175–182.
[50] T. Naito, Y. Saito, J. Yamamoto, Y. Nozaki, K. Tomura, M. Hazama, et al., Putative pheromone receptors related to the Ca2+-sensing receptor in Fugu, Proc. Natl. Acad. Sci. USA 95 (1998) 5178–5181.
[51] K. Hagino-Yamagishi, K. Moriya, H. Kubo, Y. Wakabayashi, N. Isobe, S. Saito, et al., Expression of vomeronasal receptor genes in Xenopus laevis, J. Comp. Neurol. 472 (2004) 246–256.
[52] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 25 (1997) 4876–4882.
[53] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.
[54] M. Nei, S. Kumar, Molecular Evolution and Phylogenetics, Oxford Univ., New York, 2000.
[55] J. Felsenstein, confidence limits on phylogenies: an approach using the bootstrap, Evolution 39 (1985) 783–791.
[56] A. Kumar, C.A. Dudley, R.L. Moss, Functional dichotomy within the vomeronasal system: distinct zones of neuronal activity in the accessory olfactory bulb correlate with sex-specific behaviors, J. Neurosci. 19 (1999) RC32.
[57] J. Zhang, H.F. Rosenberg, M. Nei, Positive Darwinian selection after gene duplication in primate ribonuclease genes, Proc. Natl. Acad. Sci. USA 95 (1998) 3708–3713.
[58] Z. Yang, R. Nielsen, N. Goldman, A.M. Pedersen, Codon-substitution models for heterogeneous selection pressure at amino acid sites, Genetics 155 (2000) 431–449.
[59] Z. Yang, PAML: a program package for phylogenetic analysis by maximum likelihood, Comput. Appl. Biosci. 13 (1997) 555–556.