

**EVOLUTION OF THE VOMERONASAL SYSTEM VIEWED THROUGH
SYSTEM-SPECIFIC GENES**

by

Wendy E. Grus

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in The University of Michigan
2008

Doctoral Committee:

Associate Professor Jianzhi Zhang, Chair
Professor Robert M. Bradley
Professor Priscilla K. Tucker
Associate Professor Diarmaid Ó Foighil
Associate Professor Heather L. Eisthen, Michigan State University

ACKNOWLEDGEMENTS

It would probably be easier to acknowledge everyone who didn't help me, than to acknowledge everyone who did. This list just seems so long. Alas, I will try.

First and foremost, I would like to thank my advisor Dr. Jianzhi Zhang for letting me join his lab and then giving me encouragement to be a productive member of the scientific community. I learned so much from him and I am so grateful that I could be a member of his lab.

Second, I would like to thank Dr. Soochin Cho. Soochin was always willing to give me advice on my experiments, or knew just who to ask for reagents or further advice. He has been a great friend to me in the lab, and I appreciate his kindness.

I must thank all of my other labmates, who are crazy and brilliant. I will miss the three hour lab meetings greatly. I have enjoyed being your English teacher and appreciate you putting up with my ridiculous lab decorations. I have also enjoyed being a patron of the Bakewell Bioinformatics Library and being a founding member of the Center for Dragon Genome Informatics.

I want to thank my EEB friends, especially Heather Adams and Cori Richards, for trying to keep me sane through these six years. Along those lines, I must thank the Math department 2003 cohort who has provided me with many relaxing breaks, especially playing dodgeball.

I would like to thank the EEB and GSTP faculty for encouragement and support through my stay at Michigan. I'd especially like to thank my committee: Profs. Priscilla Tucker, Diarmaid O'Foighil, Heather Eisthen, Robert Bradley, and Christopher Nosrat.

I have to thank my co-author on Chapters 3 and 4, Dr. Peng Shi. Hopefully we will be able to collaborate again in the future, when we each have our own labs.

I would like to thank Prof. Tom Finger, Dr. Anne Hansen, Nicole Shultz, Jennell Barrows, Prof. Gerry Smith, Prof. Barry O'Connor, Dr. Ron Oldfield, Dr. Prosanta Chakrabarty, Prof. Doug Futuyma, and the Hammond Bay Biological Station for help with my sea lamprey project. Additional acknowledgements go to the brave BioBev cooler for fearlessly transporting sea lampreys to Ann Arbor. I would never have been able to do it, without all the help.

I would like to thank the EEB and GSTP staff for helping my life run smoothly: LaDonna Walker, Julia Eussen, Kate Collins, Kaye Hill, Susan Stark, Jane Sullivan, David Bay, Sue Burke etc.

I would like to thank Rackham Graduate School, the EEB department, and the NIH Genome Science Training Program for funding me during these years.

Finally, I would not have made it through grad school without my hot boyfriend Mark Iwen. His love, support, and encouragement have kept me going when everything was going wrong.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	ii
LIST OF FIGURES.....	vi
LIST OF TABLES.....	viii
ABSTRACT.....	ix
INTRODUCTION.....	1
CHAPTER 1: RAPID TURNOVER AND SPECIES-SPECIFICITY OF VOMERONASAL RECEPTOR GENES IN MICE AND RATS.....	11
1.1 ABSTRACT.....	11
1.2 INTRODUCTION.....	12
1.3 MATERIALS AND METHODS.....	14
1.4 RESULTS.....	16
1.5 DISCUSSION.....	21
1.6 ACKNOWLEDGMENTS.....	26
1.7 LITERATURE CITED.....	33
CHAPTER 2: DRAMATIC VARIATION OF THE V1R VOMERONASAL RECEPTOR GENE REPERTOIRE AMONG FIVE ORDERS OF PLACENTAL AND MARSUPIAL MAMMALS.....	35
2.1 ABSTRACT.....	35
2.2 INTRODUCTION.....	36
2.3 MATERIALS AND METHODS.....	38
2.4 RESULTS.....	40
2.5 DISCUSSION.....	45
2.6 ACKNOWLEDGMENTS.....	53
2.7 LITERATURE CITED.....	61
CHAPTER 3: LARGEST VERTEBRATE VOMERONASAL TYPE 1 RECEPTOR (V1R) GENE REPERTOIRE IN THE SEMI-AQUATIC PLATYPUS.....	65
3.1 ABSTRACT.....	65
3.2 INTRODUCTION.....	65
3.3 MATERIALS AND METHODS.....	67
3.4 RESULTS AND DISCUSSION.....	68
3.5 ACKNOWLEDGMENTS.....	74
3.6 LITERATURE CITED.....	79
CHAPTER 4: DISTINCT EVOLUTIONARY PATTERNS BETWEEN CHEMORECEPTORS OF TWO VERTEBRATE OLFACTORY SYSTEMS AND THE DIFFERENTIAL TUNING HYPOTHESIS.....	82
4.1 ABSTRACT.....	82

4.2 INTRODUCTION.....	83
4.3 MATERIALS AND METHODS.....	86
4.4 RESULTS.....	90
4.5 DISCUSSION.....	94
4.6 ACKNOWLEDGMENTS.....	98
4.7 LITERATURE CITED.....	107
CHAPTER 5: ORIGIN OF THE VOMERONASAL SYSTEM IN THE COMMON ANCESTOR OF VERTEBRATES.....	111
5.1 ABSTRACT.....	111
5.2 INTRODUCTION.....	112
5.3 MATERIALS AND METHODS.....	115
5.4 RESULTS.....	117
5.4 DISCUSSION.....	122
5.5 ACKNOWLEDGMENTS.....	128
5.6 LITERATURE CITED.....	136
CONCLUSION.....	141
APPENDIX	144

LIST OF FIGURES

Figure 1.1 Unrooted phylogenetic tree of 95 putatively functional V1r genes in rat and 137 in mouse.....	27
Figure 1.2 Chromosomal locations of putatively functional rat and mouse V1r genes.	28
Figure 1.3 Synonymous (dS) and nonsynonymous (dN) distances between 18 orthologous V1r gene pairs of mouse and rat.....	29
Figure 1.4 Synonymous (dS) and nonsynonymous (dN) distances between paralogous V1r genes within families.....	30
Figure 1.5 Estimates of the dates of duplication events in the evolution of the rat V1r superfamily.....	31
Figure 2.1 Variation in the numbers of intact V1R genes (shaded bars) and pseudogenes (open bars) among mouse, rat, human, dog, cow, and opossum, with phylogenetic relationships among the species shown by a tree.....	54
Figure 2.2 Phylogeny of intact V1R genes: 8 in dog, 32 in cow, 4 in human, 102 in rat, and 187 in mouse.....	55
Figure 2.3 Phylogeny of intact V1R genes from the opossum, mouse, rat, dog, and cow.....	56
Figure 2.4 Correlation between the morphological complexity of VNO and the number of intact V1R genes in the genome.....	57
Figure 2.5 Phylogenetic positions of (A) 17 of 22 dog and (B) 35 of 41 cow V1R pseudogenes.....	58
Figure 3.1 Phylogenetic trees of vertebrate intact (A) V1Rs and (B) V2Rs.....	75
Figure 3.2 Positive correlation of vomeronasal organ complexity and V1R repertoire size.....	76
Figure 3.3 No correlation between VNO complexity and V2R repertoire size.....	77

Figure 4.1 Unrooted phylogenetic trees of all putatively functional frog, chicken, platypus, opossum, dog, and mouse nasal chemoreceptors (A) V1Rs, (B) V2Rs, (C) ORs, (D) Class I ORs, (E) Class II ORs from families 5 and 8 as defined by the HORDE database, and (F) TAARs.....	100
Figure 4.2 Proportion of genes belonging to lineage-specific clades \pm one standard error in MOS receptors (white bars) and VNS receptors (gray bars) in (A) distantly related species frog, platypus, opossum, and mouse, and (B) the closely related species mouse and rat.....	101
Figure 4.3 Ligand-binding sites are conserved in mouse-rat orthologous ORs, but not in lineage-specific paralogous mouse ORs.....	102
Figure 4.4 Pairwise amino acid sequence divergence and OR response-profile distance are significantly correlated between closely related dog OR paralogs (pluses), but not between more distantly related OR paralogs.....	103
Figure 4.5 Proportion of genes belonging to lineage-specific clades \pm one standard error in MOS receptors (white bars) and VNS receptors (gray bars) in (A) distantly related species frog, platypus, opossum, and mouse, and (B) the closely related species mouse and rat.....	104
Figure 4.6 Unrooted phylogenetic trees of all putatively functional mouse and rat nasal chemoreceptors (A) V1Rs, (B) V2Rs, (C) ORs, (D) Class I ORs, (E) Class II ORs from families 5 and 8 as defined by the HORDE database, and (F) TAARs.....	106
Figure 5.1 V1Rs are in the sea lamprey and elephant shark genomes and are expressed in the sea lamprey olfactory epithelium.....	129
Figure 5.2 V2Rs are present in the elephant shark genomes.....	130
Figure 5.3 Trpc2 is present in the sea lamprey and elephant shark genomes.....	131
Figure 5.4 Genetic components of the taste system are present in the elephant shark genome.....	133
Figure 5.5 Origins of the vertebrate chemosensory systems are staggered throughout vertebrate evolution.....	134

LIST OF TABLES

Table 1.1 V1r families in rat and mouse.....	32
Table 2.1 V1R gene family groups in five placental mammals.....	59
Table 2.2 Size variation of some gene families among mammalian species.....	60
Table 3.1 Sizes of nasal chemosensory receptor gene repertoires in vertebrates.....	78
Table 5.1 Primers for sea lamprey VNS genes.....	135

ABSTRACT

Using three genes specific to the vomeronasal system (VNS), I addressed three questions about the evolution of this vertebrate sensory system that were previously unanswerable with only morphological data. **(1)** I investigated how the V1R vomeronasal receptor repertoire evolves in mammals. **(2)** I investigated how the patterns of evolution in VNS receptors compare to those of the main olfactory system (MOS). **(3)** I investigated when the VNS originated in vertebrate evolution. For the first question, I focused on three particular aspects of mammalian V1R evolution. First, I investigated how species-specificity evolves between two closely related mammals, mouse and rat, revealing that a gene-sorting birth and death model of evolution results in many species-specific duplication and loss of V1Rs and very few orthologous mouse-rat V1R pairs. Second, I investigated V1R repertoire size variation among five orders of mammals. Dramatic variation was observed with functional repertoire size varying over 20 times between dog and mouse. This study showed a correlation between VNS morphological complexity and V1R repertoire size. Finally, I examined V1R evolution in the platypus and found that it has the largest V1R repertoire thus far identified in vertebrates. These studies revealed independent expansions in V1Rs in all three mammalian lineages. To address my second question, I compared the proportion of genes resulting from lineage-specific gene gain or loss events for nasal chemoreceptors from both vertebrate olfactory systems. With this quantitative and functional comparative study, I revealed that a significantly higher proportion of VNS receptors evolved via lineage-specific events than

did MOS receptors. The evolutionary patterns observed are consistent with the differential tuning hypothesis with main olfactory receptors being broadly tuned generalists and the VNS receptors being narrowly tuned specialists. Finally, to address my third question, I investigated the phylogenetic distribution of the VNS genetic components in early diverging lineages and determined that the VNS originated in the common ancestor of vertebrates. My work highlights the utility of system-specific genes and comparative genomics in understanding the evolution of a physiological system, and presents a much richer evolutionary history of the VNS than was thought to exist by morphological data alone.

INTRODUCTION

Being able to sense cues from the environment and from conspecifics is essential for an animal's survival. Such interactions are mediated by the animal's chemosensory systems. Understanding how these systems vary in animals that live in different environments and social structures is an interesting evolutionary question. Most vertebrates have two nasal chemosensory systems, the main olfactory system (MOS) and the vomeronasal system (VNS). The MOS is characterized morphologically by the main olfactory epithelium (MOE), sensory epithelium in the nasal cavity containing sensory neurons, and the main olfactory bulb (MOB), a region of the brain excited by the MOS sensory neurons. The VNS is characterized morphologically by the vomeronasal organ (VNO), an organ in the nasal cavity with sensory epithelium containing sensory neurons, and the accessory olfactory bulb (AOB), a region of the brain excited by the VNS sensory neurons. The MOS morphological components are found in almost all vertebrate lineages, except for some losses in marine mammals. The VNS morphological components have a more limited distribution with both morphological components appearing only in tetrapod lineages and have been lost independently in many tetrapod lineages.

Similar to their distinct morphological characters, the signal transduction pathways for the systems are also unique (Halpern and Martinez-Marcos 2003; Rodriguez 2003; Ma 2007). For the MOS, the signal is initiated by the ligand binding to

one of two types of MOS-specific G protein coupled receptors (GPCRs), odorant receptors (ORs) (Buck and Axel 1991) or trace amine associated receptors (TAARs) (Liberles and Buck 2006). This binding alters the conformation of the receptor, leading to the release of the G protein, which activates an MOS-specific adenylyate cyclase, which generates an increase of cyclic AMP (cAMP) within the neuron. The cAMP binds to a cyclic-nucleotide gated channel, which allows an influx of calcium. For the VNS, the signal is initiated by a ligand binding to one of two types of VNS-specific GPCRs, V1Rs or V2Rs. This binding alters the conformation of the receptor, leading to the release of the G protein, which activates phospholipase C (PLC). The activated PLC increases levels of two secondary messengers, diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP₃) (Minke and Cook 2002). Both DAG and IP₃ increase the intracellular calcium level; DAG by activating the Trpc2 channel allowing an inward calcium flux and IP₃ by allowing the release of intracellular calcium stores. While some of the signal transduction molecules are common to other signal transduction pathways, three types of VNS genes (V1Rs, V2Rs, and Trpc2) are known to function in the VNS-specific chemoreception.

The MOS signal is transduced in ciliated neurons found in the MOE. The axons of these neurons end at glomeruli in the MOB. Besides the sensory neurons, the MOE is composed of microvillar sustentacular supporting cells and undifferentiated basal cells. A small fraction of the olfactory sensory neurons have microvilli rather than cilia, but these neurons do not use the standard MOS signal transduction pathway and could play a role in neuronal regeneration (Elsaesser et al. 2005; Montani et al. 2006). In contrast, the VNS signal is transduced in microvillar sensory neurons in the VNO sensory epithelium.

The vomeronasal sensory neuron axons end at glomeruli in the AOB. Besides the sensory neurons and basal undifferentiated cells, ciliated cells are found in the nonsensory epithelium of mammalian VNOs and ciliated sustentacular cells are found in the VNO sensory epithelium of some amphibians (Doving and Trotier 1998). In fishes, which do not have a separate VNO, the MOE contains cells with three different sensory neuron morphologies: ciliated, microvillar, and crypt cells. It is intriguing to think of the ciliated neurons as the equivalent of the tetrapod MOS while the microvillar sensory neurons are the equivalent of the tetrapod VNS (Eisthen 1992; Eisthen 2004). As such, it would be predicted that the VNS genes are expressed exclusively in microvillar neurons. While this is generally the case, both goat and human V1Rs have expression in the MOE ciliated neurons (Rodriguez et al. 2000; Wakabayashi et al. 2002). Additionally, frog V1R expression was reported in the MOE (Date-Ito et al. 2008). However, the amphibian results are misleading in that the V1Rs were found to be expressed in the *Xenopus* olfactory chamber that is characteristically most like a teleost olfactory organ, containing both ciliated and microvillar sensory neurons. Furthermore, it is unclear whether these frog V1Rs are expressed in the ciliated neurons or the microvillar neurons of this organ (K. Hagino-Yamagishi, *personal communication*). Despite these potential exceptions, this example of finer scale sensory neuron morphology coupled to system-specific gene expression allows for further investigations of system-specific evolution.

Both the MOS and the VNS have distinct morphological and genetic components to help elucidate their distinct evolutionary histories. While many of these characters are found across vertebrates, some olfactory subsystems have been identified that are unique to certain mammalian lineages (reviewed in Breer et al 2006; Ma 2007). For example, in

some mammals, two distinct clusters of olfactory neurons in the nasal cavity are found outside of the MOE and the VNO. The first, the septal organ, (Ma et al. 2003) likely uses the same MOS signal transduction pathway to relay the signal of nine ORs that account for 95% of all septal organ OR expression (Ma 2007). The second, the Grueneberg ganglion (Fleischer et al. 2006) also expresses ORs during embryonic stages, but the majority of its sensory neurons express a unique V2R2 receptor (Fleischer et al. 2006) or TAARs (Fleischer et al. 2007). Besides these morphologically distinct subsystems, there are also molecularly distinct olfactory subsystems. For example, a small number of the sensory neurons in the main olfactory epithelium use an alternative signal transduction pathway with type D guanylyl cyclase (GC-D), which is activated by cyclic GMP rather than cAMP (see refs. in Ma 2007). The axons of these neurons project to specific regions of the main olfactory bulb known as the necklace glomeruli. Interestingly, this olfactory system has been lost in some primates, including humans, due to the pseudogenization of GC-D (Young et al. 2007). Additionally, rodent VNOs show segregated expression of vomeronasal neurons expressing either V1Rs or V2Rs. The V1R-expressing neurons coupling with $G_{\alpha_{i2}}$ G proteins are spatially distinct from the V2R-expressing neurons which couple with G_{α_0} G proteins (Halpern and Martinez-Marcos 2003). However, not all mammalian VNOs have this same segregated pattern (Takigami et al. 2004). Furthermore, rodent V2Rs selectively co-express with M10 and M1 families of MHC class 1b molecules, functioning as escort molecules in the V2R transport to the cell membranes (Ishii et al. 2003; Loconto et al. 2003). More recently it was shown that not all rodent V2Rs coexpress with these MHC molecules, resulting in a third vomeronasal subsystem (Ishii and Mombaerts 2008). Collectively, these results show the extreme

variability in vertebrate olfactory systems. However, the consistent distinction between the core genetic components of the VNS and the MOS allow for studies of the evolution of these systems based on their system-specific genes.

The identification of the nasal chemoreceptors has allowed for a deeper investigation into the evolution of the VNS. Following the discovery of ORs in the MOS (Buck and Axel 1991), two unrelated classes of GPCRs with exclusive expression in the VNS were identified, V1Rs and V2Rs (Dulac and Axel 1995; Herrada and Dulac 1997; Matsunami and Buck 1997; Ryba and Tirindelli 1997). Sequence diversity within and between species makes these sequences difficult to amplify experimentally by PCR, and complete genome sequence is required for identification of the entire repertoire for either VNS receptor family. The complete repertoire of V1Rs was available from mouse following the completion of the mouse genome sequence (Rodriguez et al. 2002), revealing a large family divided among 12 V1R subfamilies. However, having a single species V1R repertoire did not reveal much about the evolution of species specificity of V1Rs. A second mammalian V1R repertoire available, human, did not provide a useful comparison because humans do not have a VNO and likely do not have a functional VNS given that one of the major genetic components of the VNS signal transduction pathway is a pseudogene in humans (Liman and Innan 2003; Zhang and Webb 2003). Although a few human V1Rs with a complete open reading frame were identified (Rodriguez and Mombaerts 2002), the majority of these are likely relics of an ongoing pseudogenization process (Zhang and Webb 2003). Besides VNS expression, nine V1R genes were found to be expressed in the testis and hypothesized to play a role in sperm maturation or migration (Tatsura et al. 2001), but mice deficient for these V1Rs are fertile (Del Punta

et. al 2002). Two V1R knockout studies (Boschat et al. 2002; Del Punta et al. 2002) demonstrate that at least some V1Rs are pheromone receptors, but they do not exclude the possibility that other V1Rs can detect non-pheromonal chemicals. In any case, because the functions of V1R receptors are VNS-specific, the evolution of mammalian V1Rs can tell us the evolution of the mammalian VNS.

In the first part of my thesis, I focus on the evolution of V1Rs in mammals. In Chapter 1, I present a more meaningful comparison to understand the evolution of mammalian V1Rs by comparing the identified mouse V1Rs with V1Rs mined from the rat genome, as rats are known to have a functional VNS, V1Rs were originally described from rats (Dulac and Axel 1995), and mouse and rat are closely related species that share orthologs for the majority of their genes (Rat Genome Sequencing Consortium 2004). I find that species-specificity is manifested through gene-sorting evolution of V1Rs whereby genes are randomly duplicated and lost. Both ancient and recent duplication events and ancient gene loss have resulted in mouse and rat V1R repertoires with a surprisingly low number of orthologous gene pairs. Based on the identification of the mouse, rat, and human V1R repertoires, it would be expected that all mammals have a large V1R repertoire; even humans, without a functional VNS, have a large repertoire of V1R pseudogenes. To determine if the large V1R repertoire size is characteristic of all mammals, despite known variation in VNO morphology, in Chapter 2, I identified the V1R repertoire from cow, dog, and opossum and compare those to the already described mouse, rat, and human V1Rs. Surprisingly, cow and dog have much smaller V1R repertoires, suggesting that there is similar variability between VNO complexity and V1R repertoire size. Because platypus has the most complex type of mammalian VNO

(Wysocki 1979), in Chapter 3, I identified the platypus V1R repertoire to see if the correlation between VNO complexity and V1R repertoire size held even for this divergent mammalian lineage. Platypus has the largest V1R repertoire size so far examined. These results are viewed in light of other platypus sensory systems.

In the second part of my thesis, I use available VNS receptors and MOS receptors to compare evolutionary patterns and potential functional differences between the two systems. For a comparative analysis of the MOS and VNS, reciprocally unique genetic characters must also be known for both systems. In Chapter 4, I quantitatively compare the VNS receptors and MOS receptors from mouse, rat, dog, opossum, platypus, chicken, and frog. The VNS receptors for mouse, rat, opossum, platypus, and frog have a significantly higher proportion of genes from lineage-specific gene gain/loss events than do the MOS receptors. These results coupled to analysis of putative ligand binding sites in orthologous and paralogous ORs support the differential tuning hypothesis whereby MOS receptors are broadly tuned generalist receptors and the VNS receptors are narrowly tuned specialist receptors.

In the final part of my thesis, I use VNS-specific genes to determine the origin of the VNS in vertebrates. Based on the phylogenetic distribution of the morphological components, the VNS was thought to have evolved as a terrestrial adaptation (Bertmar 1981). However, further fine-scale examination, such as sensory neuron morphology, suggests that the VNS might exist in an unrecognized form in teleost fish (Eisthen 1992), but a different VNS marker was needed to determine the presence of a precursor to the VNS (Eisthen 2004). Based on the phylogenetic distribution of the VNS-specific genetic components and their expression patterns in non-tetrapod lineages, the VNS transduction

pathway originated to before the common ancestor of teleost fish and tetrapods (Grus and Zhang 2006). In Chapter 5, I investigate whether the three VNS-specific genes are present in early diverging vertebrate lineages, jawless fish and cartilaginous fish. I find two and three VNS-specific genes in each of these lineages, respectively, suggesting that the VNS arose in the common ancestor of extant vertebrates.

LITERATURE CITED

- Bertmar, G. 1981. Evolution of vomeronasal organs in vertebrates. *Evolution* **35**:359-366.
- Boschat, C., C. Pelofi, O. Randin, D. Roppolo, C. Luscher, M. C. Broillet, and I. Rodriguez. 2002. Pheromone detection mediated by a V1r vomeronasal receptor. *Nat Neurosci* **5**:1261-1262.
- Breer, H., J. Fleischer, and J. Strotmann. 2006. The sense of smell: multiple olfactory subsystems. *Cell Mol Life Sci* **63**:1465-1475.
- Buck, L., and R. Axel. 1991. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**:175-187.
- Date-Ito, A., H. Ohara, M. Ichikawa, Y. Mori, and K. Hagino-Yamagishi. 2008. Xenopus VIR Vomeronasal Receptor Family Is Expressed in the Main Olfactory System. *Chem Senses* **33**:339-346.
- Del Punta, K., T. Leinders-Zufall, I. Rodriguez, D. Jukam, C. J. Wysocki, S. Ogawa, F. Zufall, and P. Mombaerts. 2002. Deficient pheromone responses in mice lacking a cluster of vomeronasal receptor genes. *Nature* **419**:70-74.
- Doving, K. B., and D. Trotier. 1998. Structure and function of the vomeronasal organ. *J Exp Biol* **201**:2913-2925.
- Dulac, C., and R. Axel. 1995. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83**:195-206.
- Eisthen, H. L. 2004. The goldfish knows: olfactory receptor cell morphology predicts receptor gene expression. *J Comp Neurol* **477**:341-346.
- Eisthen, H. L. 1992. Phylogeny of the vomeronasal system and of receptor cell types in the olfactory and vomeronasal epithelia of vertebrates. *Microsc Res Tech* **23**:1-21.
- Elsaesser, R., G. Montani, R. Tirindelli, and J. Paysan. 2005. Phosphatidyl-inositide signalling proteins in a novel class of sensory cells in the mammalian olfactory epithelium. *Eur J Neurosci* **21**:2692-2700.
- Fleischer, J., K. Schwarzenbacher, S. Besser, N. Hass, and H. Breer. 2006. Olfactory receptors and signalling elements in the Grueneberg ganglion. *J Neurochem* **98**:543-554.
- Fleischer, J., K. Schwarzenbacher, and H. Breer. 2007. Expression of trace amine-associated receptors in the Grueneberg ganglion. *Chem Senses* **32**:623-631.
- Grus, W. E., and J. Zhang. 2006. Origin and evolution of the vertebrate vomeronasal system viewed through system-specific genes. *Bioessays* **28**:709-718.
- Halpern, M., and A. Martinez-Marcos. 2003. Structure and function of the vomeronasal system: an update. *Prog Neurobiol* **70**:245-318.
- Herrada, G., and C. Dulac. 1997. A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. *Cell* **90**:763-773.
- Ishii, T., J. Hirota, and P. Mombaerts. 2003. Combinatorial coexpression of neural and immune multigene families in mouse vomeronasal sensory neurons. *Curr Biol* **13**:394-400.
- Ishii, T., and P. Mombaerts. 2008. Expression of nonclassical class I major histocompatibility genes defines a tripartite organization of the mouse vomeronasal system. *J Neurosci* **28**:2332-2341.

- Liberles, S. D., and L. B. Buck. 2006. A second class of chemosensory receptors in the olfactory epithelium. *Nature* **442**:645-650.
- Liman, E. R., and H. Innan. 2003. Relaxed selective pressure on an essential component of pheromone transduction in primate evolution. *Proc Natl Acad Sci U S A* **100**:3328-3332.
- Loconto, J., F. Papes, E. Chang, L. Stowers, E. P. Jones, T. Takada, A. Kumanovics, K. Fischer Lindahl, and C. Dulac. 2003. Functional expression of murine V2R pheromone receptors involves selective association with the M10 and M1 families of MHC class Ib molecules. *Cell* **112**:607-618.
- Ma, M. 2007. Encoding olfactory signals via multiple chemosensory systems. *Crit Rev Biochem Mol Biol* **42**:463-480.
- Ma, M., X. Grosmaître, C. L. Iwema, H. Baker, C. A. Greer, and G. M. Shepherd. 2003. Olfactory signal transduction in the mouse septal organ. *J Neurosci* **23**:317-324.
- Matsunami, H., and L. B. Buck. 1997. A multigene family encoding a diverse array of putative pheromone receptors in mammals. *Cell* **90**:775-784.
- Minke, B., and B. Cook. 2002. TRP channel proteins and signal transduction. *Physiol Rev* **82**:429-472.
- Montani, G., S. Tonelli, R. Elsaesser, J. Paysan, and R. Tirindelli. 2006. Neuropeptide Y in the olfactory microvillar cells. *Eur J Neurosci* **24**:20-24.
- Rat Genome Sequencing Consortium. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**:493-521.
- Rodriguez, I. 2003. Nosing into pheromone detectors. *Nat Neurosci* **6**:438-440.
- Rodriguez, I., K. Del Punta, A. Rothman, T. Ishii, and P. Mombaerts. 2002. Multiple new and isolated families within the mouse superfamily of V1r vomeronasal receptors. *Nat Neurosci* **5**:134-140.
- Rodriguez, I., C. A. Greer, M. Y. Mok, and P. Mombaerts. 2000. A putative pheromone receptor gene expressed in human olfactory mucosa. *Nat Genet* **26**:18-19.
- Rodriguez, I., and P. Mombaerts. 2002. Novel human vomeronasal receptor-like genes reveal species-specific families. *Curr Biol* **12**:R409-411.
- Ryba, N. J., and R. Tirindelli. 1997. A new multigene family of putative pheromone receptors. *Neuron* **19**:371-379.
- Takigami, S., Y. Mori, Y. Tanioka, and M. Ichikawa. 2004. Morphological evidence for two types of Mammalian vomeronasal system. *Chem Senses* **29**:301-310.
- Tatsura, H., H. Nagao, A. Tamada, S. Sasaki, K. Kohri, and K. Mori. 2001. Developing germ cells in mouse testis express pheromone receptors. *FEBS Lett* **488**:139-144.
- Wakabayashi, Y., Y. Mori, M. Ichikawa, K. Yazaki, and K. Hagino-Yamagishi. 2002. A putative pheromone receptor gene is expressed in two distinct olfactory organs in goats. *Chem Senses* **27**:207-213.
- Wysocki, C. J. 1979. Neurobehavioral evidence for the involvement of the vomeronasal system in mammalian reproduction. *Neurosci Biobehav Rev* **3**:301-341.
- Young, J. M., H. Waters, C. Dong, H. J. Fulle, and E. R. Liman. 2007. Degeneration of the olfactory guanylyl cyclase D gene during primate evolution. *PLoS ONE* **2**:e884.
- Zhang, J., and D. M. Webb. 2003. Evolutionary deterioration of the vomeronasal pheromone transduction pathway in catarrhine primates. *Proc Natl Acad Sci U S A* **100**:8337-8341.

CHAPTER 1

RAPID TURNOVER AND SPECIES-SPECIFICITY OF VOMERONASAL RECEPTOR GENES IN MICE AND RATS

1.1 ABSTRACT

Pheromones are used by individuals of the same species to elicit behavioral or physiological changes, and they are perceived by the vomeronasal organ (VNO) in some terrestrial vertebrates. Vomeronasal receptors are encoded by the V1r and V2r gene superfamilies in mammals. A comparison of the V1r and V2r repertoires between closely related species can provide significant insights into the evolutionary genetic mechanisms responsible for species-specific chemosensory communications. 137 putatively functional V1r genes of 12 families were previously identified from the mouse genome. We here report the identification of 95 putatively functional V1r genes from the draft rat genome sequence. These genes map primarily to four blocks in two chromosomes. The rat V1r genes can be phylogenetically grouped into 10 families that are shared with mouse and two new families that are rat-specific. Even in many shared families gene numbers differ between the two species, apparently due to frequent gene duplication and pseudogenization after the separation of the two species. Molecular dating suggests that most of the rat V1r families emerged before or during the radiation of mammalian orders, but many duplications within families occurred as recently as in the past 10 MY. Our results show that the evolution of the V1r repertoire is characterized by exceptionally fast

gene turnover via gains and losses of individual genes, suggesting rapid and substantial changes in chemosensory communication between species.

1.2 INTRODUCTION

Pheromones are used by individuals of the same species to elicit behavioral or physiological changes such as male-male aggression, puberty, estrus, and induction of mating, and are perceived to varying degrees by the vomeronasal organ (VNO) in some mammals (reviewed in Keverne 1999). Two superfamilies of vomeronasal receptors, V1r and V2r, are known in mammals and they differ in expression location and gene structure (Dulac and Axel 1995; Herrada and Dulac 1997; Matsunami and Buck 1997; Ryba and Tirindelli 1997). While both types of receptors are seven-transmembrane G-protein coupled receptors, V1rs are characterized by an intronless coding region, while V2rs are characterized by a long highly variable N-terminal domain. V1rs are expressed in G α 2 neurons and V2rs are expressed in G α 0 neurons (reviewed in Dulac and Torello, 2003). Targeted deletion of some V1r genes in mice show altered aggression and sexual behaviors (Del Punta et al., 2002). Additionally, a third vomeronasal receptor superfamily, V3r, has been described (Pantages and Dulac 2000); however, V3rs were later found to be a family of V1rs (Rodriguez et al. 2002). Because they lack introns in the coding region, V1r genes are more accessible than V2r genes to bioinformatic as well as experimental studies. For example, the entire mouse V1r repertoire, including 137 putatively functional members of 12 families has been described (Rodriguez et al. 2002), while the mouse V2r repertoire has yet to be described.

The identification of V1r and V2r genes has opened the door to studies on the molecular mechanisms and origins of species-specific communications. V1rs were originally discovered and described in rat (Dulac and Axel 1995). The absence of highly conserved regions in V1r prohibits the design of degenerate primers that can amplify a large number of genes across wide taxonomic scale (Giorgi and Rouquier 2002). Therefore, the V1r superfamily was not extensively described in any species until the availability of the human and mouse genome sequences (Lane et al. 2002; Rodriguez et al. 2002; Rodriguez and Mombaerts 2002). The comparison of V1rs between human and mouse is not informative because the two species are distantly related and because humans lack VNO sensitivity due to loss of important components of the VNO signal transduction pathway (Liman and Innan 2003; Zhang and Webb 2003). Zhang et al. (2004) compared the V1r repertoires of two genome assemblies generated from different strains of inbred mice. While this comparison is useful for identifying polymorphisms within species, it does not address interspecific differences, which are the hallmark of pheromone communication. The availability of the draft rat genome sequence (Rat Genome Sequencing Consortium 2004) provides the opportunity to compare for the first time vomeronasal receptor repertoires of relatively closely related species, as mouse and rat diverged only 18 ± 6 million years (MY) ago (Rat Genome Sequencing Consortium, 2004).

Divergence in the V1r repertoires between species can occur in three ways: functional divergence of orthologs, loss or gain of family members, and loss or gain of entire families. Two recent studies compared a small number of V1r genes between mouse and rat (Lane et al., 2004; Emes et al., 2004) but did not provide a general picture

on the evolution of the V1r repertoire. We here compare the entire V1r repertoires in an attempt to examine which of the three processes dominates the divergence of V1r genes between species.

1.3 MATERIALS AND METHODS

1.3.1 Database searches

TBLASTN searches for rat V1r genes were done on the rat genome sequence available in the National Center for Biotechnology Information (NCBI) with the rat genome build 2 version 1 (<http://www.ncbi.nlm.nih.gov/genome/guide/rat/index.html>). The previously described 137 mouse V1r genes were used as query sequences. V1r pseudogenes were identified by premature stop codons or incomplete sequence across the 13 protein domains (7 transmembrane, 3 extracellular, and 3 intracellular) of the 15 domains. The N-terminal extracellular and C-terminal intracellular domains were not considered in our criterion because they are highly variable in sequence length. The physical locations of the rat genes were determined by mapping the TBLASTN results to chromosomal contigs. The rat genes were named by their family memberships (e.g., V1re5) and the numbers after the family designation are randomly chosen.

1.3.2 Sequence alignment and phylogenetic analysis

Gene sequences were aligned according to the protein sequence alignment made by ClustalX (Thompson et al. 1997) with manual adjustment. Phylogenetic trees were constructed using the neighbor-joining method (Saitou and Nei 1987) with 2000 bootstrap replications (Felsenstein 1985). Synonymous (d_S) and nonsynonymous (d_N)

substitution rates were computed by the modified Nei-Gojobori method (Zhang et al. 1998). MEGA2 (Kumar et al., 2001) was used for these evolutionary analyses.

1.3.3 Identifying putative rat-mouse orthologs

To find orthologous V1r genes, we identified clades in the phylogenetic tree that contain a single rat gene and a single mouse gene and then calculated the d_S between the putative orthologs. Those with estimated d_S within the range of 0.063 to 0.317 were accepted as orthologs. This range covers 95% of over 11,000 orthologous mouse-rat gene pairs (Rat Genome Sequencing Consortium, 2004). In the present study, 19 putative orthologous pairs were identified via the phylogenetic analysis and 18 of them passed the above d_S criterion. Our procedure is conservative in the sense that it is unlikely to generate false orthologous pairs but may miss some real pairs.

1.3.4 Dating duplication events

To determine the dates of duplication events within the rat V1r families, we calculated the pairwise d_S for paralogous genes within families. The average synonymous substitution rate in rodents is about $0.19/36=5.3\times 10^{-9}$ per site per year, estimated using the mean d_S of 0.19 and the divergence time of 18 ± 6 million years between the mouse and rat (Rat Genome Sequencing Consortium, 2004).

To determine the time of duplication between gene families, we calculated the average d_S between families and used the above calibration. We assumed that the root of the tree in Fig. 1.1 is on the stem branch leading to V1rc, as suggested by Rodriguez et al. (2002) when taste receptor genes were used to root the mouse V1r tree.

1.4 RESULTS

1.4.1 Composition of the rat V1r gene superfamily

From the database searches with mouse V1r queries, we identified from the rat genome sequence 116 V1r genes. Additional searches with human V1r-like genes (Rodriguez and Mombaerts, 2002) did not yield any additional sequences. Since the rat genome is over 90% complete (Rat Genome Sequencing Consortium, 2004), the V1r repertoire we describe here is probably over 90% complete as well. We examined the conceptually translated protein sequences encoded by the retrieved DNA sequences and found that 95 of them contain open reading frames (ORFs) that cover 13 internal domains. These 95 ORFs are regarded as putatively functional genes. It is possible that some of them may in fact be nonfunctional due to mutations in either protein-coding or regulatory regions. The proportion of putatively functional V1r genes appears higher in rats (95/116=82%) than in mice (53%; Rodriguez et al., 2002; Zhang et al., 2004), although this may be caused by the bias of TBLASTN against pseudogene detection.

A neighbor-joining tree of all putatively functional V1r genes from the mouse and rat was reconstructed using protein sequences (Fig. 1.1). The 137 mouse V1r genes were previously classified into 12 families based on phylogenetic analysis as well as the criterion that genes with >40% protein sequence identity belong to the same family (Rodriguez et al., 2002). These 12 families were recovered in our tree of mouse and rat genes, and most (10/12) of them have high bootstrap support (>90%). In 10 (a, b, c, d, e, f, g, j, k, and l) of the 12 families, rat genes are also found (Table 1.1), and these are the shared V1r families between the two species. Two mouse families (h and i) do not

contain any rat genes and are mouse-specific families. In addition, two new families (m and n; defined by the same criterion as used in mouse) contain only rat genes (Table 1.1) and are referred to as rat-specific families. Because the rat genome sequence was obtained from an inbred line (Rat Genome Sequencing Consortium, 2004) and the V1r genes were identified by their chromosomal location, it is unlikely that genes with high sequence identity are only allelic variants and not separate genes. In fact, two of the rat V1rd genes (V1rd6 and V1rd7) are identical in DNA sequence, but located 1 Mb apart on chromosome 1. Our evolutionary analysis provides further support that V3r (V1rd) genes (Pantages and Dulac, 2000) are not a separate vomeronasal receptor superfamily (Rodriguez et al., 2002) but are a family of V1r genes (Fig. 1.1). In fact, only eight V1rd genes are present in the rat genome.

1.4.2 Chromosomal organization of V1r genes

The 95 putatively functional rat V1r genes map to 8 locations in chromosomes 1, 4, and 7, although most of them are located in four major blocks in chromosomes 1 and 4 (Fig. 1.2). It is interesting to note that members of the same family tend to be located in the same chromosomal region, suggesting tandem gene duplication as the primary mechanism for family expansion.

Using the mouse-rat homologous chromosome map (<http://www.genboree.org>), we found syntenic regions between the four major blocks of V1r genes in rat and four blocks in mouse (Fig. 1.2). Both blocks on rat chromosome 4 map to mouse chromosome 6. The two blocks on rat chromosome 1 are split mainly between mouse chromosomes 7 and 17. Twenty-eight putatively functional mouse V1r genes identified by Rodriguez et al. (2002) have yet to be mapped.

From the phylogeny shown in Fig. 1.1 and estimates of synonymous distances, we identified 18 orthologous V1r gene pairs between mouse and rat. These pairs are shown on the chromosomes as well (Fig. 1.2). One V1rg pair (rat V1rg9 and mouse V1rg10) has a mouse gene in chromosome 5, not part of the large block in V1rg family in mouse (Fig. 1.2). However, a comparative study of a different mouse strain did not identify any V1r genes on mouse chromosome 5 (Zhang et al., 2004), indicating that this is a recent translocation within the species *Mus musculus*.

1.4.3 Comparison with mouse V1r genes

The most notable difference between the V1r repertoires of the mouse and rat is the presence of two mouse-specific and two rat-specific families. The mouse-specific V1rh and V1ri families contain 23% of all mouse V1r genes. We determined that the absence of rat V1rh and V1ri genes is not due to lack of sequencing in this region, as we can find rat orthologs of non-V1r genes that are located in the mouse V1rh and V1ri chromosomal regions and a V1ri pseudogene was found in this region in rat. The rat-specific V1rm and V1rn families contain 6% of all rat V1r genes. No functional genes or pseudogenes that belong to these two families were found in mouse.

Fig. 1.1 shows that in addition to the species-specific gene families, many shared families have different numbers of genes in the two species. V1rc is the richest family in both species. However, V1re contains 21 rat genes but only 13 mouse genes. On the other hand, V1rd has 8 rat but 22 mouse genes. These differences show that lineage-specific contraction and/or expansion of certain V1r families must have occurred after the mouse-rat separation. In some shared V1r families (e.g., V1ra), not a single one-to-one

orthologous pair is found between the two species. This implies that every gene in this family has been subject to either duplication or pseudogenization since the mouse-rat split.

1.4.4 d_N/d_S ratio for orthologous genes and paralogous genes

To examine whether positive selection has been operating during the divergence of orthologous V1r genes between species, we computed the number of synonymous substitution per synonymous site (d_S) and the number of nonsynonymous substitution per nonsynonymous site (d_N) for the 18 pairs of orthologous genes between mouse and rat. A d_N/d_S ratio that significantly exceeds 1 provides the most convincing evidence for positive selection, whereas $d_N/d_S < 1$ shows overall purifying selection on the gene, although positive selection at a small number of sites cannot be excluded.

The d_S values for the 18 V1r gene pairs range from 0.14 to 0.26 with a mean of 0.19, identical to the mean d_S from over 11,000 orthologous gene pairs between the mouse and rat (Rat Genome Sequencing Consortium, 2004). All 18 d_S values fall within the middle 67% of the d_S distribution determined from this large set of orthologous genes (Rat Genome Sequencing Consortium, 2004), suggesting that our determination of orthology was correct.

The overall d_N/d_S ratio across the entire gene sequence is less than 1 for each of the 18 pairs of orthologous genes (Fig. 1.3A), with an average d_N/d_S of 0.46. This value is in the 2% upper tail of the distribution of d_N/d_S derived from the over 11,000 gene comparisons aforementioned (mean=0.11; Rat Genome Sequencing Consortium, 2004). As mean d_S for the 11,000 genes and the mean d_S for the V1r genes are virtually

identical, the difference in the ratio comes from a much higher d_N in the V1r genes than in most other genes. This may be due to relatively weak purifying selection on V1r genes or the presence of positive selection at some sites. However, even when we examined d_N/d_S for separate domains, all show $d_N/d_S < 1$ (Fig. 1.3B).

We also compared members of the same V1r families in rats to test whether positive selection may be responsible for divergence of V1r duplicates within families. The overall d_N/d_S ratio averages 0.549. The d_N/d_S ratios from the rat specific families (Fig. 1.4A) and the mouse-specific families (Fig. 1.4B) are similar to the d_N/d_S ratios from the families shared between both species (Fig. 1.4C, 1.4D). We did not attempt to estimate d_N/d_S ratios between genes belonging to different V1r families because of the high sequence divergence and expected low reliability of estimation caused by multiple hits.

1.4.5 Dates of V1r gene duplications

It is important to know the approximate dates of gene duplication events that gave rise to the V1r genes. Because d_S increases over time relatively constantly, we estimated d_S values between paralogous V1r genes of rats. A calibration of $d_S=0.19$ corresponding to a gene age of 18 MY was used and dates of the duplication events within and between rat V1r families were obtained (Fig. 1.5A, 1.5C). Although many duplication events within gene families occurred after the mouse-rat split, about half of them also took place before this split. Additionally, one-fourth of all the duplications that occurred within families are estimated to have taken place in the last 10 MY, including very recent duplications within V1rd, V1re, V1rf, and V1rm. It appears that the number of

duplications that led to stably retained genes increased in the recent past, as indicated by a negative correlation ($r^2=0.75$, $P<0.003$) between the number of duplications and evolutionary age (Fig. 1.5B). However, this could be due to short lifespan for V1r genes, such that ancient duplicates are more likely to have been lost than young ones, rendering underestimation of duplication numbers in the ancient past.

Molecular dating suggests that all of the rat V1r families were present at the time of split between rat and mouse (Fig. 1.5C). Even the youngest rat V1r family, the one descended from the common ancestor of V1ra and V1rb, emerged ~62 MY ago. These estimates are conservative because large d_s values are likely to be underestimated due to the difficulty of correcting multiple hits at synonymous sites.

1.5 DISCUSSION

In this study, we identified 95 putatively functional V1r genes from the rat genome sequence, supporting the prediction of ~100 distinct genes in the original description of the rat V1r gene superfamily (Dulac and Axel, 1995). This number is about two thirds of that in the mouse. Wild mice and rats have different social/reproductive structures (Abbott 2004). Mice live in groups with one highly aggressive alpha male monopolizing the females, whereas rats are promiscuous and less aggressive. It is possible that the observed difference in the number of their V1r genes has biological significance.

Our study localized the majority of the rat V1r genes to two blocks on chromosome 4 and two blocks on chromosome 1, which show synteny to four chromosomal regions in mouse. While the families making up each region are the same,

there are few orthologous gene pairs within them. The Rat Genome Sequencing Project Consortium (2004) estimated that 86-94% of rat genes have one-to-one mouse orthologs. However, only 19% of the rat V1r genes have one-to-one mouse orthologs, indicative of rapid gene turnover. Indeed, many V1r families differ in size between the two species and each species has two families that are absent in the other species. All these observations show frequent gains and losses of V1r genes during evolution. Such a pattern of gene family evolution has often been observed in host-defense genes such as the Major-Histocompatibility-Complex, immunoglobulin, and eosinophil-associated RNase genes (Cadavid et al. 1997; Sitnikova and Su 1998; Zhang et al. 2000). While the rapid turnover of host-defense genes is presumably a response to ever-changing pathogens that infect hosts, that of V1r genes could be due to rapid change in pheromone or odorant cues during evolution, which may be under intense sexual and/or natural selection.

The gene family tree in Fig. 1.1 gives the impression that the V1r repertoire enlarges over evolutionary time. This interpretation of the tree could be wrong. It is quite possible that the V1r repertoire in the common ancestor of mice and rats was as complex as those of today's mice and rats. Since new genes gained in evolution show up in the tree whereas old genes that have been lost no longer appear, the tree looks as if there are more genes now than in the past. In other words, one should not interpret the evolutionary pattern of V1r genes as gene family expansion (Lane et al. 2002) unless there is other evidence. A more appropriate term is "gene sorting", which describes unequal loss, retention, and amplification of ancestral genes between species (Zhang et al. 2000) and is consistent with the birth-and-death model of gene family evolution (Nei

et al., 1997). For example, molecular dating suggests that the rat-specific V1rm and V1rn families were present in the common ancestor of the mouse and rat. Thus, the absence of these two families in the mouse must be due to lineage-specific loss that occurred in the mouse lineage since it was separated from the rat lineage. Similarly, the absence of the mouse-specific V1rh and V1ri families in rat must be due to the lineage-specific loss in rats. In fact, a V1ri pseudogene is still present in the rat genome.

In addition to the gains and losses of entire V1r families and gains and losses of V1r genes within families, divergence between mouse and rat V1r repertoires may also occur by functional divergence of orthologous genes. At present, there is no direct evidence showing functional differences between mouse and rat orthologous V1r genes. If positive selection is found to act in the divergence of orthologous genes, one may infer that there is functional divergence. This is because positive selection occurs only if there is a functional change that results in increased organismal fitness. In our analysis, we did not find evidence for higher d_N over d_S for any of the 18 orthologous gene pairs compared. However, the average d_N/d_S for the V1r genes is significantly higher than that for other mouse/rat genes, indicating that V1rs are subject to either weak purifying selection or positive selection that may act at only a few sites of the protein. In this regard, it is interesting to note that a few earlier studies have suggested action of positive selection on V1rs. Emes et al. (2004) analyzed 22 mouse and rat V1r genes by a likelihood method (Yang et al., 2000) and identified positive selection acting on 14 codons. Lane et al. (2004) used the same statistical method in an analysis of 14 mouse and rat V1r genes and identified positive selection at 5-10% of codons. Since most of their sequences were paralogous, it is not possible to infer whether there has been

positive selection between orthologs. Furthermore, the statistical method they used has been shown to be unreliable because it often detects positive selection when there is none (Suzuki and Nei 2001; 2002; 2004). Recently Zhang et al. (2004) conducted a comprehensive comparison of V1r genes detected from two mouse genome assemblies. Since the two assemblies were derived from different inbred strains of mice, the orthologous differences observed reflect intraspecific polymorphisms. These authors found a d_N/d_S ratio of 1.13 at the polymorphic sites. Although this was presented as evidence for positive selection, we think that it should be interpreted with care for three reasons. First, it has not been shown that the d_N/d_S ratio is significantly higher than 1, thus the observation does not reject the neutral hypothesis. In fact, a relatively high d_N/d_S ratio within species could result from ineffective purifying selection if there has been recent population shrinking. Second, the observed between-strain genetic differences were likely present in mouse populations before breeding, as the history of managed breeding is short. However, because of intense artificial selection and/or extensive genetic drift during breeding, the differences between the two inbred strains may not accurately represent the polymorphic pattern in natural mouse populations. Third, the evolutionary patterns of V1r genes may be different at the intra- and inter-specific levels because odorants potentially mediate species-specific communications. Thus, even if positive selection is acting at the intra-specific level, it may not act at the inter-specific level. Taken together, although it is quite likely that orthologous V1r genes between species have been under positive selection, no conclusive evidence is available.

Previous studies of limited numbers of mouse and rat V1rs suggested that many V1r gene duplications occurred at the time of split between mouse and rat. Our

comprehensive analysis of all rat V1rs do not show a particularly high number of duplication events around that time. Instead, the number of successful duplications appears to be higher in recent times than in the past (Fig. 1.5). This observation, if coupled with a constant number of gene loss per unit time, would suggest that the V1r repertoire is expanding. It will be interesting to estimate the gene loss rate using pseudogene data as well as functional gene data from additional species that are closely related to the mouse and rat. Our mapping data suggest that gains of V1r genes often occur by tandem gene duplication. Lane and colleagues (2004) found that these duplications appear to be mediated by repetitive elements in the genome.

Our molecular dating of the V1r gene families, although approximate, provides two interesting results. First, the V1r families of the mouse and rat were established early in mammalian evolution. The majority of the families observed in mouse and rat were likely present in the most recent common ancestor of extant rodents, which existed about 75 MY ago (Springer et al., 2003). The second observation is that many V1r families emerged between 90 and 140 MY ago. This is the time shortly before and during the radiation of placental mammals (80-110 MY ago; Springer et al., 2003). Although entirely speculative, it is tempting to hypothesize that the rapid diversification of V1r families enabled the development of advanced and intricate vomeronasal-mediated communications, which facilitated increased speciation and mammalian radiation. Characterizing V1r and V2r genes from more placental mammals as well as from marsupials would help test this idea.

1.6 Acknowledgments

We thank Soochin Cho and David Webb for valuable comments. This work was supported by a startup fund from the University of Michigan and National Institutes of Health grant GM67030 to J.Z.

Figure 1.1 Unrooted phylogenetic tree of 95 putatively functional V1r genes in rat and 137 in mouse. Rat genes are identified with the suffix R. Families V1ra-V1rl were previously defined by Rodriguez et al. (2002) and families V1rm and V1rn are newly identified here. Bootstrap percentages supporting the monophyly of each family are given, except for V1ra, which is not monophyletic due to one sequence (V1ra9). The tree was reconstructed using the neighbor-joining method with Poisson-corrected protein distances. The scale bar shows 0.1 amino acid substitution per site.

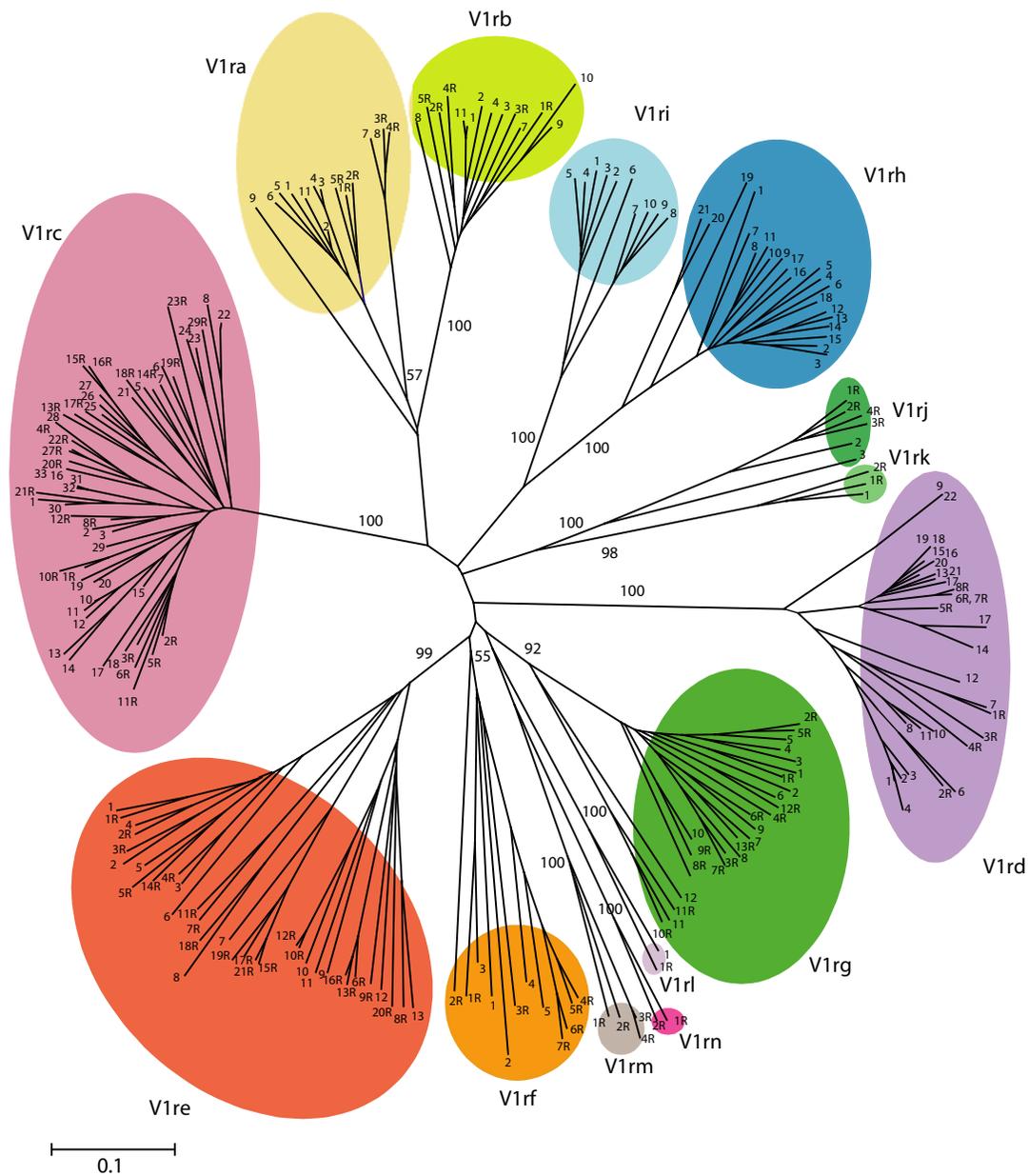


Figure 1.2 Chromosomal locations of putatively functional rat and mouse V1r genes. Lines between mouse and rat genes indicate orthologous relationships. Unmapped mouse genes are shown as “UN”. Colors designate family membership as in Fig. 1.1.

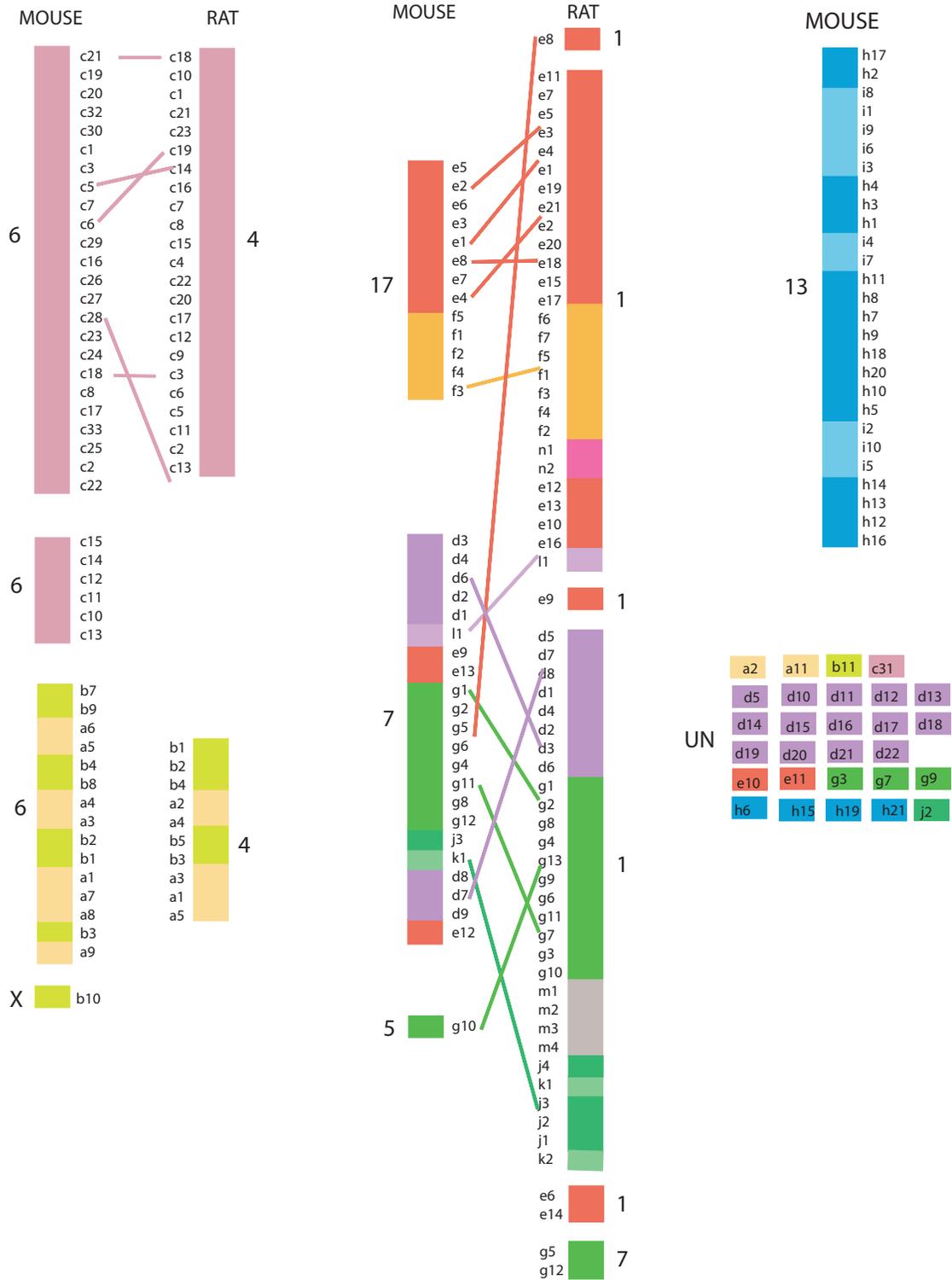


Figure 1.3 Synonymous (dS) and nonsynonymous (dN) distances between 18 orthologous V1r gene pairs of mouse and rat. (A) Overall dS and dN across the entire sequences. The dotted vertical line indicates $dS = 0.19$, which is the mean dS for over 11,000 mouse-rat orthologous gene pairs analyzed by Rat Genome Sequencing Consortium (2004). The diagonal line shows $dN = dS$. (B) The average dN (black bars) and dS (white bars) among the 18 orthologous pairs for each of the 13 domains. TM, transmembrane domain; EC, extracellular domain; IC, intracellular domain. The N-terminal extracellular domain and C-terminal intracellular domain are not studied due to high variation in sequence length. Error bars show one standard error of the mean. The dashed line is the average dS for the entire gene, whereas the dotted line is the average dN for the entire gene.

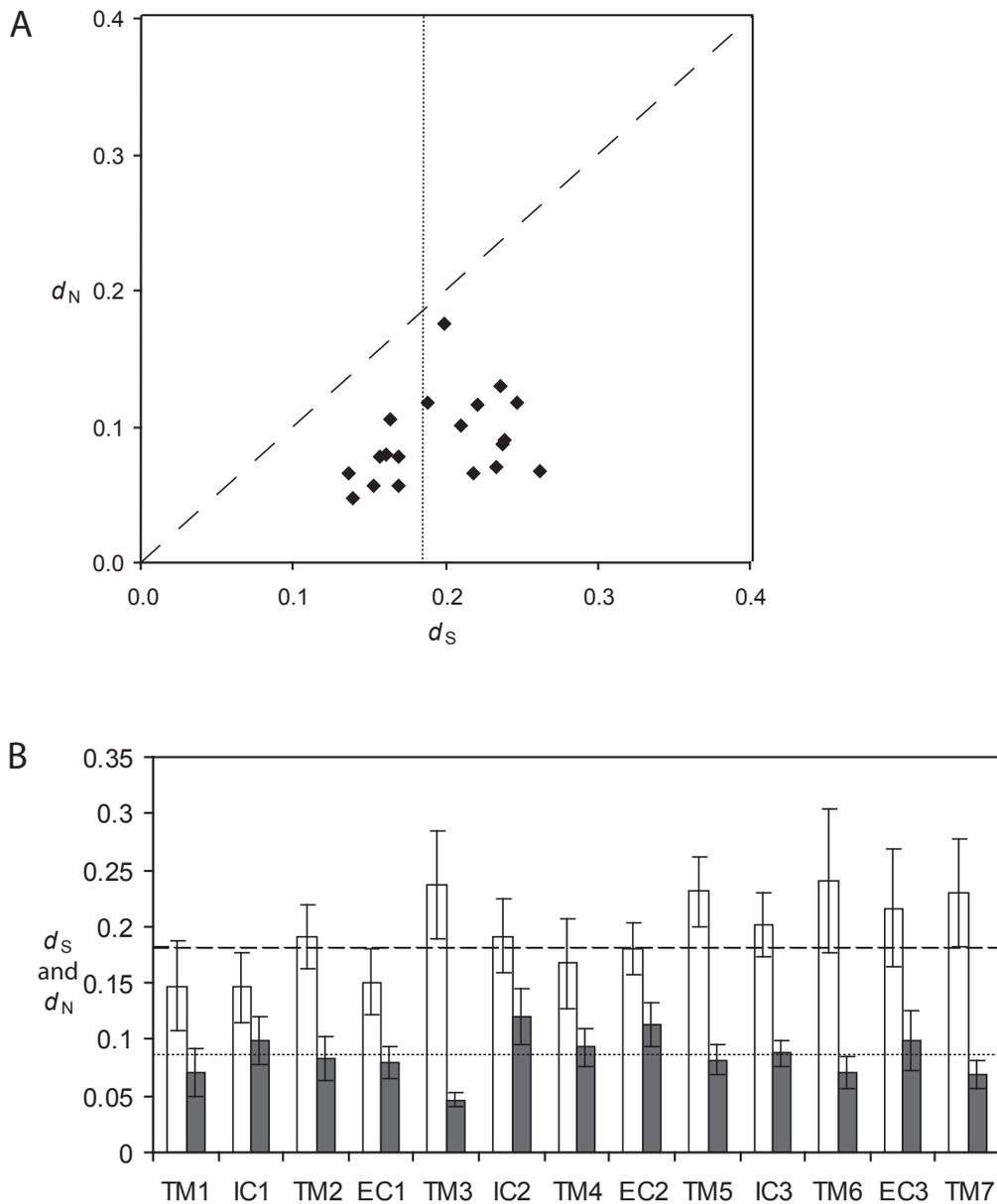


Figure 1.4 Synonymous (d_S) and nonsynonymous (d_N) distances between paralogous V1r genes within families. The diagonal lines show $d_N=d_S$. (A) d_S and d_N for rat-specific families. Black and white symbols represent V1rm and V1rn, respectively. (B) d_S and d_N for mouse-specific families. Black and white symbols represent V1rh and V1ri, respectively. (C) d_S and d_N between paralogous rat genes in V1ra, a typical V1r family shared between mouse and rat. (D) d_S and d_N between paralogous mouse genes in V1ra.

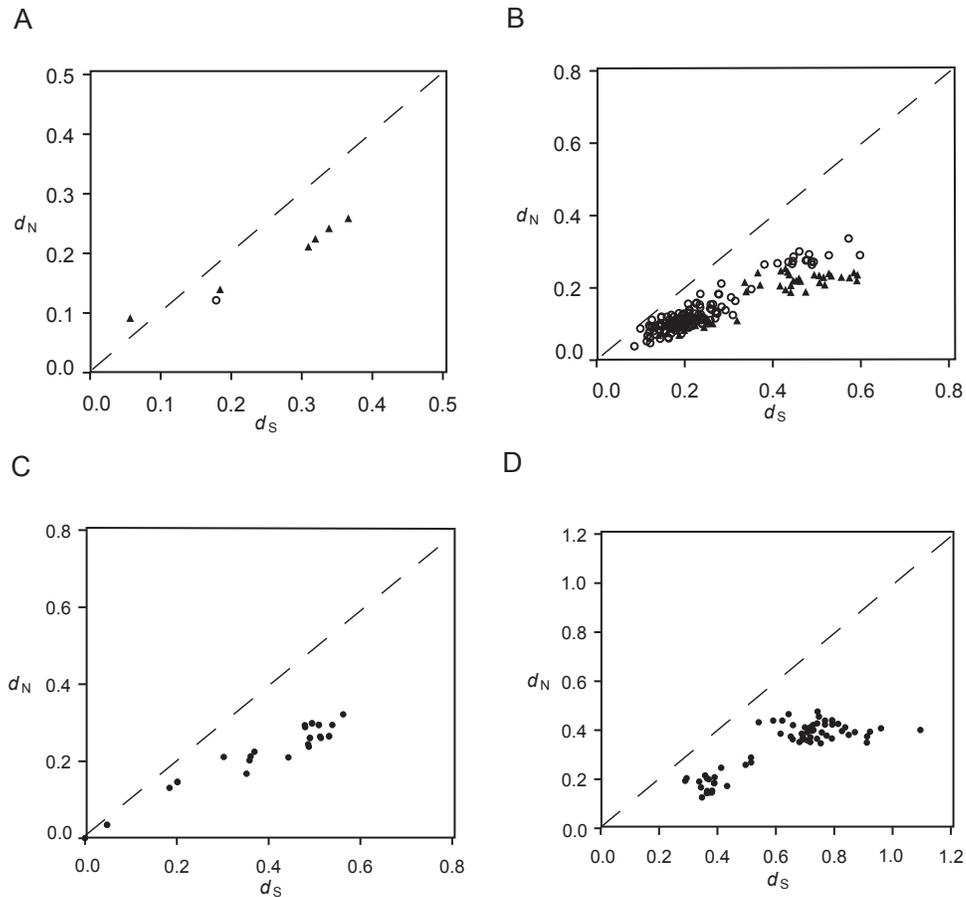


Figure 1.5 Estimates of the dates of duplication events in the evolution of the rat V1r superfamily. (A) Dates for duplications within families. Each dot represents a duplication event. The vertical dashed line represents the time when mouse and rat were separated. The vertical dotted lines divide the figure into 10 MY intervals. (B) The distribution of the number of duplication events in 10 MY intervals. (C) The dates of the duplication events leading to the 12 families found in rat. The dashed line indicates the time of rat-mouse divergence.

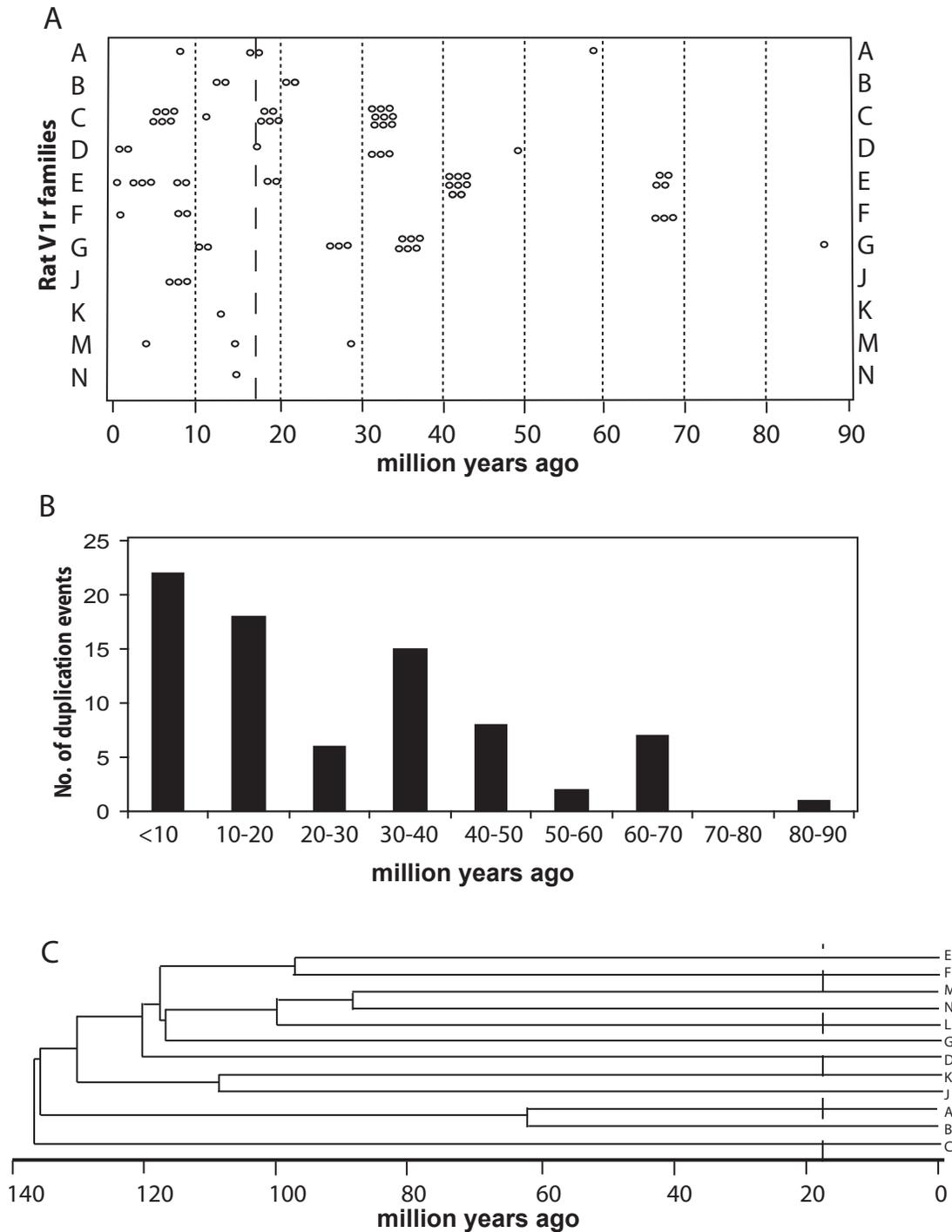


Table 1.1: V1r families in rat and mouse

Family	No. of functional genes		% identity (amino acid)	
	Rat	Mouse	Rat	Mouse
A	5	10	57-92	54-99
B	5	9	72-84	70-98
C	23	31	74-96	60-99
D	8	22	54-100	61-93
E	21	13	43-97	44-77
F	7	5	43-93	45-75
G	13	12	48-91	53-88
H	0	21	--	61-93
I	0	10	--	65-91
J	4	2	84-90	58
K	2	1	84	--
L	1	1	--	--
M	4	0	70-97	--
N	2	0	81	--
Total	95	137		

1.7 LITERATURE CITED

- Abbott, A. 2004. Laboratory animals: the Renaissance rat. *Nature* **428**: 464-466.
- Cadavid, L.F., C. Shufflebotham, F.J. Ruiz, M. Yeager, A.L. Hughes, and D.I. Watkins. 1997. Evolutionary instability of the major histocompatibility complex class I loci in New World primates. *Proc Natl Acad Sci U S A* **94**: 14536-14541.
- Dulac, C. and R. Axel. 1995. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83**: 195-206.
- Emes, R.D., S.A. Beatson, C.P. Ponting, and L. Goodstadt. 2004. Evolution and comparative genomics of odorant- and pheromone-associated genes in rodents. *Genome Res* **14**: 591-602.
- Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783-791.
- Giorgi, D. and S. Rouquier. 2002. Identification of V1R-like putative pheromone receptor sequences in non-human primates. Characterization of V1R pseudogenes in marmoset, a primate species that possesses an intact vomeronasal organ. *Chem Senses* **27**: 529-537.
- Herrada, G. and C. Dulac. 1997. A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. *Cell* **90**: 763-773.
- Keverne, E.B. 1999. The vomeronasal organ. *Science* **286**: 716-720.
- Lane, R.P., T. Cutforth, R. Axel, L. Hood, and B.J. Trask. 2002. Sequence analysis of mouse vomeronasal receptor gene clusters reveals common promoter motifs and a history of recent expansion. *Proc Natl Acad Sci U S A* **99**: 291-296.
- Lane, R.P., J. Young, T. Newman, and B.J. Trask. 2004. Species specificity in rodent pheromone receptor repertoires. *Genome Res* **14**: 603-608.
- Liman, E.R. and H. Innan. 2003. Relaxed selective pressure on an essential component of pheromone transduction in primate evolution. *Proc Natl Acad Sci U S A* **100**: 3328-3332.
- Matsunami, H. and L.B. Buck. 1997. A multigene family encoding a diverse array of putative pheromone receptors in mammals. *Cell* **90**: 775-784.
- Pantages, E. and C. Dulac. 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., K. Del Punta, A. Rothman, T. Ishii, and P. Mombaerts. 2002. Multiple new and isolated families within the mouse superfamily of V1r vomeronasal receptors. *Nat Neurosci* **5**: 134-140.
- Rodriguez, I. and P. Mombaerts. 2002. Novel human vomeronasal receptor-like genes reveal species-specific families. *Curr Biol* **12**: R409-411.
- Ryba, N.J. and R. Tirindelli. 1997. A new multigene family of putative pheromone receptors. *Neuron* **19**: 371-379.
- Saitou, N. and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406-425.
- Sitnikova, T. and C. Su. 1998. Coevolution of immunoglobulin heavy- and light-chain variable-region gene families. *Mol Biol Evol* **15**: 617-625.

- Suzuki, Y. and M. Nei. 2001. Reliabilities of parsimony-based and likelihood-based methods for detecting positive selection at single amino acid sites. *Mol Biol Evol* **18**: 2179-2185.
- Suzuki, Y. and M. Nei. 2002. Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. *Mol Biol Evol* **19**: 1865-1869.
- Suzuki, Y. and M. Nei. 2004. False-Positive Selection Identified by ML-Based Methods: Examples from the Sig1 Gene of the Diatom *Thalassiosira weissflogii* and the tax Gene of a Human T-cell Lymphotropic Virus. *Mol Biol Evol* **21**: 914-921.
- Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876-4882.
- Zhang, J., K.D. Dyer, and H.F. Rosenberg. 2000. Evolution of the rodent eosinophil-associated RNase gene family by rapid gene sorting and positive selection. *Proc Natl Acad Sci U S A* **97**: 4701-4706.
- Zhang, J., H.F. Rosenberg, and M. Nei. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci U S A* **95**: 3708-3713.
- Zhang, J. and D.M. Webb. 2003. Evolutionary deterioration of the vomeronasal pheromone transduction pathway in catarrhine primates. *Proc Natl Acad Sci U S A* **100**: 8337-8341.
- Zhang, X., I. Rodriguez, P. Mombaerts, and S. Firestein. 2004. Odorant and vomeronasal receptor genes in two mouse genome assemblies. *Genomics* **83**: 802-811.

CHAPTER 2

DRAMATIC VARIATION OF THE V1R VOMERONASAL RECEPTOR REPERTOIRE AMONG FIVE ORDERS OF PLACENTAL AND MARSUPIAL MAMMALS

2.1 ABSTRACT

Pheromones are chemicals emitted and sensed by conspecifics to elicit social and sexual responses, and are detected in terrestrial vertebrates by the vomeronasal organ (VNO). Vomeronasal receptors are encoded by the V1R and V2R vomeronasal gene superfamilies. The V1R superfamily contains 187 and 102 putatively functional genes in the mouse and rat, respectively. To investigate whether this large repertoire size is typical among mammals with functional VNOs, we here describe the V1R repertoires of dog, cow, and opossum, based on their draft genome sequences. The dog and cow have only 8 and 32 intact V1R genes, respectively. Thus, the intact V1R repertoire size varies by at least 23 fold among placental mammals with functional VNOs. To our knowledge, this size ratio represents the greatest among-species variation in gene family size of all mammalian gene families. Phylogenetic analysis of placental V1R genes suggests multiple losses of ancestral genes in carnivores and artiodactyls and gains of many new genes by gene duplication in rodents, manifesting massive gene births and deaths. We also identify 49 intact opossum V1R genes and discover independent expansions of the repertoire in placentals and marsupials. We further show a concordance between the V1R repertoire size and the complexity of VNO

morphology, suggesting that the latter could indicate the sophistication of chemosensory communications within species. In sum, our results demonstrate tremendous diversity and rapid evolution of mammalian V1R gene inventories and caution the generalization of VNO biology from rodents to all mammals.

2.2 INTRODUCTION

Pheromones provide conspecific chemical communications that elicit sexual and social changes in behavior and physiology (Keverne 1999). For instance, pheromonal peptides found in mouse urine can mediate individual recognition, induce early puberty, or block pregnancy (Keverne 1999), and pheromones can control behaviors such as maternal aggression (Del Punta et al. 2002). In mammals, some pheromones are sensed by the vomeronasal organ (VNO), which resides on the bottom of the nasal cavity and is anatomically and physiologically separated from the main olfactory system that detects odorants (Keverne 1999). In rodents, two superfamilies of 7-transmembrane G protein-coupled receptors, V1Rs and V2Rs, serve as vomeronasal receptors, some of which have pheromone receptor function (Dulac and Torello 2003). Individual V1R genes are encoded by an intronless region of ~1000 nucleotides and are expressed in vomeronasal sensory neurons whose cell bodies are located in the apical part of the VNO epithelium (Dulac and Axel 1995). These cells also express a G-protein subunit named $G\alpha_{i2}$ (Dulac and Torello 2003). In contrast, the multi-exon V2R genes are characterized by a long, highly variable N-terminal domain and are co-expressed with $G\alpha_o$ in sensory neurons whose cell bodies are basally located (Herrada and Dulac 1997; Matsunami and Buck 1997; Ryba and Tirindelli 1997). Both receptor superfamilies were originally identified

in rat and each of them was estimated to contain ~100 genes (Dulac and Axel 1995; Herrada and Dulac 1997; Matsunami and Buck 1997; Ryba and Tirindelli 1997). Due to the simpler gene structure, the complete repertoire of the V1R superfamily has been described in mice, rats, and humans (Grus and Zhang 2004; Rodriguez et al. 2002; Rodriguez and Mombaerts 2002; Shi et al. 2005; Zhang et al. 2004). The superfamily has 187 and 102 functional genes in the mouse and rat, respectively (Grus and Zhang 2004; Shi et al. 2005). In humans, ~200 V1R sequences have been identified, although only four of them have intact open reading frames (ORFs) in the majority of individuals (Rodriguez and Mombaerts 2002; Zhang and Webb 2003). The massive V1R pseudogenization observed in the human genome started shortly before the separation of hominoids (i.e., humans and apes) from Old World monkeys, apparently because of the reduced importance of vomeronasal-mediated communications, and the four human V1R ORFs are likely relics of this ongoing pseudogenization process (Zhang and Webb 2003).

The morphological complexity of the VNO as well as the complexity of VNO-mediated olfactory communications varies substantially among different mammals (Takami 2002). These variations led to the hypothesis that the complexity of the V1R and V2R repertoires may also vary greatly among species. However, this hypothesis has not been rigorously tested, because the absence of highly conserved regions in V1Rs makes it difficult to design degenerate primers to amplify a large number of genes across a wide taxonomic scale (Giorgi and Rouquier 2002; Mombaerts 2004). Although a few limited studies in selected mammals suggested possible among-species variation in the numbers of V1R genes and pseudogenes (Giorgi and Rouquier 2002; Wakabayashi et al. 2002), the extent of this variation and the exact numbers of genes and pseudogenes in

these species are unknown. Furthermore, the long-term evolutionary pattern of V1R genes in mammals and other vertebrates remains largely uncharacterized. With the recent availability of the draft genome sequences of dog, cow, and opossum, we now describe the complete V1R repertoires of this diverse array of species. We report a variation in the size of the V1R repertoire among five orders of placental and marsupial mammals that is surprisingly large and unprecedented in any other mammalian gene family.

2.3 METHODS

2.3.1 Database searches

TBLASTN searches for V1R genes were conducted on the dog (*Canis familiaris*), cow (*Bos taurus*), and opossum (*Monodelphis domestica*) genome sequences. The 7.6X coverage dog genome sequence (<http://www.ncbi.nlm.nih.gov/genome/guide/dog/>) and 3.3X coverage cow genome sequence (<http://www.ncbi.nlm.nih.gov/genome/guide/cow/>) are available in the National Center for Biotechnology Information (NCBI). The 7.2X coverage opossum genome sequence is available at ENSEMBL (http://pre.ensembl.org/Monodelphis_domestica/). Previously published mouse and rat V1R genes (Grus and Zhang 2004; Rodriguez et al. 2002; Shi et al. 2005) were used as query sequences. Putative V1Rs were identified with an E-value cutoff of 10^{-5} . They were then used as queries to BLAST the nr database of GenBank. A putative V1R gene was considered to be real if its best hit was a previously known V1R. Use of human V1Rs (Rodriguez and Mombaerts 2002) as query sequences did not yield additional V1Rs. V1Rs were considered pseudogenes if they contained premature stop codons or

were incomplete across the 13 internal domains (7 transmembrane, 3 extracellular, and 3 intracellular). The database searches were independently conducted by the first two authors, yielding identical results. The sequences of newly identified intact V1Rs from the dog, cow, and opossum can be found online (<http://www.pnas.org/cgi/data/0501589102/DC1/1>).

2.3.2 Sequence alignment and phylogenetic analysis

Gene sequences were aligned per protein sequence alignment by Clustal_X (Thompson et al. 1997) with manual adjustment. Phylogenetic trees were reconstructed using the neighbor-joining method (Saitou and Nei 1987) with protein-Poisson distances (Nei and Kumar 2000), and were evaluated by 1000 bootstrap replications (Felsenstein 1985). We also used protein-*p* distances (Nei and Kumar 2000) and found that the branching patterns with high bootstrap support in the Poisson-distance tree remained unchanged in the *p*-distance tree. Following convention (Rodriguez et al. 2002), we defined gene families by a minimum of 40% amino acid identity between all family members and confirmed the monophyly of gene families by phylogenetic analysis. We conceptually translated each pseudogene sequence according to its alignment with all functional V1R genes (mouse, rat, and dog or cow), and determined the phylogenetic position of each pseudogene by making a neighbor-joining tree with the protein sequence of the pseudogene and those of all functional genes. The distribution of LINE repetitive elements was determined by the RepeatMasker program (<http://www.repeatmasker.org/>). For easy comparison, we used the same criterion (Lane et al. 2002) to calculate the L1 density. More specifically, we estimated the L1 density in

the intergenic regions for tandem linked V1R genes and that in 10,000 nucleotides upstream and downstream of the coding regions for non-tandem linked V1R genes.

2.4 RESULTS

2.4.1 Dramatic variation in V1R repertoire size in placental mammals

Placental mammals can be classified into four superorders (Murphy et al. 2004). Rodents and primates belong to the superorder Euarchontoglires, to which all previous V1R evolutionary analyses have been restricted. Here we investigate the V1R repertoire in the dog and cow, which are members of the superorder Laurasiatheria, the sister clade to Euarchontoglires. From the dog genome sequence, we identified 8 complete V1R ORFs, which are presumably functional genes. In addition, 22 V1R pseudogenes were detected. Because the dog genome sequence has a high (7.6X) coverage, it is likely that the majority, if not all dog V1R genes have been found. Similarly, we detected 32 functional V1R genes and 41 pseudogenes from the cow genome sequence. However, since the cow sequence has a relatively low coverage (3.3X), it is possible that a few additional genes and pseudogenes may be discovered when a more complete genome sequence becomes available. Even with this limitation, our observations clearly show that the V1R repertoires in cow and dog are substantially smaller than those in mouse and rat, which contain 187 and 102 putatively functional genes, respectively (Shi et al. 2005). Even when humans are disregarded, there is still a 23-fold variation in the size of functional V1R repertoire among rodents, artiodactyls, and carnivores (Fig. 2.1). The proportion of putatively functional V1R genes is lower in dogs (27%) than in rodents

(49%, Rodriguez et al. 2002; Zhang et al. 2004), but the proportion in cows (44%) is similar to that in rodents.

2.4.2 Phylogenetic relationships of placental V1R genes

To understand the evolutionary relationships of the V1R genes from the four placental orders (rodents, primates, artiodactyls, and carnivores), we reconstructed a protein neighbor-joining tree with all putatively functional V1Rs from the mouse, rat, human, dog, and cow (Fig. 2.2). We also included 4 human V1Rs with complete ORFs in the phylogenetic analysis, although these ORFs are likely relics of an ongoing pseudogenization process (Zhang and Webb 2003). A fifth human V1R gene, hV1RL3, was not included because a previous study found this gene to be nearly fixed with a nonfunctional allele in human populations (allele frequency >98%) (Zhang and Webb 2003). Putatively functional non-human primate V1R sequences from (Mundy and Cook 2003) were not included because only partial sequences were available.

The dog and cow V1R genes cluster within the previously described rodent V1R superfamily (Grus and Zhang 2004). The mouse and rat V1R genes were previously classified into 14 families (V1RA to V1RN) in (Grus and Zhang 2004). A new family (V1RO) is found when the 7 recently described rat V1R genes (Shi et al. 2005) are added (see Fig. 2.2). Since the families were originally described in rodents, some of the families could be rodent-specific, meaning that their origins postdated the origin of rodents (Grus and Zhang 2004; Murphy et al. 2004). To classify other mammalian V1Rs, we clustered the families into family groups. The family groups contain families, described in rodents, that split around the time of the most recent common ancestor of

rodents, cow, and dog. The 15 rodent families may be grouped into 9 family groups based on phylogeny (Fig. 2.2, Table 2.1). Based on the rodent synonymous substitution rate, we previously estimated that the V1R family groups appeared prior to 95 MY ago (Grus and Zhang 2004). This suggests that the most recent common ancestor of primates, rodents, artiodactyls, and carnivores should have the V1R family groups that are currently observed in rodents, as this ancestor lived ~95 MY ago (Murphy et al. 2004; Springer et al. 2003). Note that this estimate for the age of the V1R family groups is conservative because it is likely that the synonymous substitution rate has been enhanced in rodents compared with that in other mammals (Li 1997; Rat Genome Sequencing Consortium 2004). We found that all nine family groups contain putatively functional mouse genes and eight of nine (except V1RH/I) contain functional rat genes (Fig. 2.2, Table 2.1). However, only five of the nine family groups include functional dog genes and five include functional cow genes (Fig. 2.2), suggesting that cows and dogs lost many ancestral V1R genes. For instance, the family groups V1RC and V1RG each contain more than ten mouse genes and ten rat genes but neither contains functional dog or cow V1Rs. Additionally, there are no functional dog V1RD or V1RE genes and there are no functional cow V1RH/I genes. Surprisingly, the second largest V1R gene family in cows (V1RD) contains no functional dog genes. We also identified two cow-specific families, V1RP and V1RQ, each containing a single gene.

Interestingly, several family groups that do not contain functional cow or dog genes possess their pseudogenes. We were able to classify 17 of the 22 dog V1R pseudogenes and 35 of 41 cow V1R pseudogenes by family group (Fig. 2.5, Table 2.1). The remaining pseudogenes have degenerated too much to be included. Family groups

V1RC and V1RD each contain two dog pseudogenes, but no functional dog genes. When both functional genes and pseudogenes are considered, only two family groups in dog (V1RE and V1RG) and two in cow (V1RC and V1RG) have been completely lost if the unclassified pseudogenes do not belong to these family groups. These observations provide further evidence that the small sizes of the dog and the cow V1R repertoires can be partially explained by loss of ancestral genes.

The second factor causing the repertoire-size variation among species is the lineage-specific expansion of families as occurred most prominently in rodents. The phylogenetic analysis shows that most of the dog and cow V1Rs diverged from their closest rodent homologs before the expansions of rodent V1R families (Fig. 2.2). This divergence pattern is consistent with our previous estimate that the earliest duplication events within rodent V1R families took place ~88 million years (MY) ago (Grus and Zhang 2004), postdating the separation of rodents, carnivores, and artiodactyls ~95 MY ago (Murphy et al. 2004; Springer et al. 2003). Family expansions were virtually absent in the dog, but were evident in three V1R families of cow. Only two dog families, V1RF and V1RJ/K, contain more than one V1R gene. The other three dog V1Rs are single genes in V1RL/M/N, V1RH, and V1RA/B/O, respectively. Six of the 22 dog pseudogenes were part of the two larger dog V1R families. In contrast, three cow V1R families, V1RD, V1RE, and V1RF, exhibited the duplications characteristic of the rodent V1R gene families. The remaining cow V1Rs are in V1RA/B/O, V1RJ/K, V1RL/M/N, V1RP, and V1RQ. While some of these families contain more than one cow gene (Table 2.1), the multiple genes are not the product of species-specific duplication events. Twenty-eight of the 41 cow pseudogenes were part of the three largest cow families. The

dog-cow V1RF clade was similar to the rodent gene families with species-specific duplications (Fig. 2.2).

Based solely on comparison with the mouse, Rodriguez and Mombaerts (2002) reported that human V1R genes do not belong to the reported V1R families. Our analysis showed that three human V1Rs can be classified into V1RF and V1RL/M/N, and the fourth (hV1RL5) forms a new human-specific family (V1RR). Giorgi and Rouquier (2002) identified several V1R sequences from the chimpanzee, gorilla, and orangutan. We found that these sequences cluster closely with the human sequences, although we did not present them in the phylogeny of Fig. 2.2 because they are not from complete V1R repertoires. Thus, when all four placental orders are considered, there are 12 V1R family groups, 10 of which have moderate to high bootstrap support (Fig. 2.2). In the case of V1RF, we maintained the family by at least 40% amino acid identity among all genes. Fig. 2.2 also shows V1RF as paraphyletic; however, the bootstrap values for the deep branches defining V1RF are low, indicating that it could be monophyletic.

2.4.3 Opossum V1R repertoire

We also identified 49 putatively functional V1R genes and 53 pseudogenes from the opossum genome sequence. Because the opossum genome sequence has a high coverage (7.2X), we expect that almost all opossum V1R genes have been detected. We reconstructed a phylogenetic tree of the 49 opossum V1Rs with all functional V1Rs of the mouse, rat, dog, and cow (Fig. 2.3). The phylogeny shows that the opossum genes can be classified into eight opossum-specific families (oV1RA to oV1RH). The families range in size from a single gene (oV1RH) to 15 genes (oV1RA and oV1RC), with

variable levels of bootstrap support. The tree shows that the placental and marsupial genes do not form two separate monophyletic groups, suggesting that more than one V1R gene was present in the common ancestor of placentals and marsupials. However, it is difficult to estimate the number of V1R genes in the common ancestor due to the low resolution of deep nodes in the tree. Nevertheless, the presence of many well-supported opossum-specific and placental-specific gene clusters in the tree provides strong phylogenetic evidence that V1R families expanded independently in marsupials and placentals.

2.5 DISCUSSION

In this study, we described the vomeronasal receptor V1R gene superfamily from the dog, cow, and opossum, extended the study of the superfamily outside of rodents and primates, and revealed extremely high variation in the sizes of functional V1R repertoires among mammals. The sizes of the dog and cow V1R repertoires are vastly smaller than those of rodents and primates. We only found 8 putatively functional V1Rs in dog and 32 in cow, compared with 102 in rat and 187 in mouse (Grus and Zhang 2004; Rodriguez et al. 2002; Shi et al. 2005; Zhang et al. 2004). In humans, the functional V1R repertoire is also small with only 4 ORFs in most individuals (Rodriguez and Mombaerts 2002; Zhang and Webb 2003). However, the entire human V1R repertoire, including both functional genes and pseudogenes, is quite large, with ~200 members. In opossum, we identify 49 putatively functional V1Rs, intermediate among what we identify here for dog and cow and what had been previously identified for rodents.

Even when humans which have lost functional VNOs are disregarded, the size of the functional V1R repertoire varies by more than 23 fold among all mammals or among placental mammals (Table 2.2). Several gene families, particularly those involved in sensory, immune, and reproductive functions, are known to vary substantially in size among mammalian species (Rat Genome Sequencing Consortium 2004). For instance, the number of functional olfactory receptor genes in rat is ~4 times that in humans (Niimura and Nei 2003; Rat Genome Sequencing Consortium 2004). The putatively anti-parasitic eosinophil-associated RNase gene family is 6-17 times larger in rodents than in New World monkeys (Cho et al. 2005; Rosenberg et al. 1995; Zhang et al. 2000). The human X-linked testis-expressed homeobox genes *OTEX* and *PEPP2* have a total of 15 orthologous genes in the mouse genome, due to multiple gene duplications that postdated the primate-rodent separation (Wang and Zhang 2004). The human genome contains over 200 immunoglobulin heavy chain variable region (VH) genes and ~80 of them are functional (Lefranc 2001). Rabbit also contains over 100 VH genes, but only one of them is predominantly used, resulting in very few functional genes (Su and Nei 1999). The exact number of functional rabbit VH genes is yet to be determined, although at least 5 have been identified (Su and Nei 1999). Thus, the number of functional VH genes may vary up to 16 fold among different mammals. Table 2.2 lists additional gene families known to have wide variations of family size among mammalian species. However, to our knowledge, the size variation in V1R repertoire among mammals exceeds that in any other mammalian gene family. This high variation might be in part because V1Rs are involved in both sensory and reproductive functions. Our phylogenetic analysis indicates that the dramatic size difference in the V1R repertoires of placental mammals is due to

two molecular evolutionary mechanisms. First, some ancestral gene families that are still present in rodents have been lost in dogs and cows. Second, species-specific duplication events characteristic of rodent V1R families were less frequent in cows and dogs. Thus, massive gene deaths and births (Hughes and Nei 1989; Nei and Hughes 1992) in different lineages explain the observed size variation.

Is it possible that the smaller V1R repertoire in dogs and cows indicates that the VNO is not functional in these organisms? Pseudogenization of vomeronasal genes and loss of VNO function happened in catarrhine primates (i.e., humans, apes, and Old World monkeys), presumably following the acquisition of full trichromatic vision (Liman and Innan 2003; Webb et al. 2004; Zhang and Webb 2003). It is possible that stereoscopic vision in both primates and carnivores (Kral 2003) compensates for reduced vomeronasal-mediated communication. Complete loss of VNO function, however, is unlikely to be responsible for the small V1R repertoires of dog and cow, because *Trp2*, the ion channel necessary for VNO signal transduction (Leypold et al. 2002; Stowers et al. 2002) is apparently functional in cow (Wissenbach et al. 1998), and we were able to identify a complete ORF for dog *Trp2* from the genome sequence. Furthermore, there have been reports of bovine pheromones that induce estrus, which are possibly mediated by the VNO (Lane and Wathes 1998; Rekwot et al. 2001). Is it possible that the V1R repertoires have shrunk during the domestications of dog and cow due to either artificial selection or genetic drift? We think it is unlikely because the domestication started no earlier than 15,000 years ago for dogs and 10,000 years ago for cow (Bradley et al. 1996; Savolainen et al. 2002). Even if a small number of functional genes have become nonfunctional during domestication, their relics should remain readily detectable as

pseudogenes. We thus believe that the sizes of V1R repertoires in dog and cow should be very close to those in their wild ancestors. Another possibility is that the small V1R repertoire could be compensated by a large V2R repertoire. This explanation also seems unlikely, as all V2Rs we identified from dog and cow were pseudogenes (data not shown). Furthermore, all V2R genes identified from the human and goat genomes are pseudogenes (Wakabayashi et al. 2002). In fact, no functional V2Rs have been reported in non-rodent mammals. If the estimated size of ~100 genes in the rodent V2R repertoire (Herrada and Dulac 1997) is accurate, the phylogenetic surveys suggest an even more dramatic variation in the V2R repertoire among placental mammals. In this respect, it is interesting to note that a recent study found two types of vomeronasal systems in mammals, with rodents and opossums having both $G\alpha_{i2}$ - and $G\alpha_O$ -expressing vomeronasal sensory neurons and all other species examined (goat, dog, horse, musk shrew, and marmoset) having only $G\alpha_{i2}$ -expressing vomeronasal sensory neurons (but see Dennis et al. 2003; Takigami et al. 2004). Because V1Rs are expressed in $G\alpha_{i2}$ positive neurons and V2Rs are expressed in $G\alpha_O$ positive neurons (Mombaerts 2004), it is possible that functional V2Rs only exist in rodents and opossums among mammals. Indeed, our preliminary search confirms the presence of V2R ORFs in the opossum genome (P. Shi and J. Zhang, unpublished results).

It should be noted that not all VNO-mediated functions in rodents are VNO-mediated in other mammals (Dennis et al. 2003) and it is possible that during the evolution of rodents some olfactory cues became detectable by the VNO. In fact, a goat V1R gene is known to be expressed in the olfactory epithelium (Wakabayashi et al. 2002). Before we can determine the effects of gene loss in V1R superfamily evolution,

we must understand the functional difference between family groups. Two studies have investigated the functions of VIRs, one in a strain of mice mutant for *VIRb2* (Boschat et al. 2002) and one in a strain of mice which had 16 genes of the families VIRA and VIRB knocked out (Del Punta et al. 2002). Both studies showed that the mutations resulted in physiological or behavioral changes in the mice. But, no study has compared the phenotypic effects of eliminating different VIR families, and it remains unknown whether each family is necessary for a specific function. Such information would allow us to reason why certain families are lost in some species.

In addition to gene loss, our analysis clearly demonstrated VIR family expansions in some placental mammals. However, it is unknown what factors have promoted the dramatic expansions in rodents but have only allowed limited expansion in cows and virtually no expansion in dogs. Lane and colleagues (2002; 2004) suggested that the VIR gene duplications in rodents were mediated by L1 repetitive elements. These elements, which are known to have been active in rodents around the time of the mouse-rat divergence, densely populate the regions of the mouse and rat genomes where VIRs are located. L1 density differs between species, so it would be interesting to see how VIR repertoire size correlates with L1 density. Following the procedure of Lane et al. (Lane et al. 2002), we found that 21% of the DNA sequences in the genomic regions harboring dog VIR genes are L1 elements, lower than the corresponding density (40%) in the mouse VIR loci (Lane et al. 2002) or the average density (25%) in the mouse genome (Smit 1999). The L1 density of the genomic regions containing cow VIR genes is also 21%. Thus, the low duplicability of both dog and cow VIR genes might be in part due to the low density of L1 elements in the genomic regions. L1 density in cow VIR

genomic regions is similar to that in dog, but cow have four times as many V1Rs as do dogs. Thus, L1 elements might not play as great a role in V1R duplications as originally thought, or the role of L1 elements in V1R duplication might be limited to rodents.

The phylogenetic reconstruction of placental V1Rs with the opossum V1Rs (Fig. 2.3) suggests that there was more than one V1R gene in the common ancestor of marsupials and placentals, and the mammalian V1R families are then at least 170 million years old (Murphy et al. 2004). Many V1R families expanded in placentals and marsupials independently. Sequencing another marsupial genome, which is currently under way for the tammar wallaby *Macropus eugenii*, will significantly broaden our understanding of V1R evolution in marsupials. Because both primates and rodents have over 100 V1R genes (or pseudogenes in the case of humans) and because primates and rodents are more closely related to each other than either of them is to carnivores or artiodactyls (Murphy et al. 2004; Springer et al. 2003), one might infer that after the separation of the common ancestor of cows and dogs from the common ancestor of rodents and primates, there was a dramatic expansion of the V1R repertoires. Thus, the large size of the V1R superfamily as observed in rodents might be restricted to organisms derived from the common ancestor of primates and rodents, including the five orders of Rodentia, Lagomorpha (e.g., rabbit), Dermoptera (e.g., flying lemur), Scandentia (e.g., tree shrew), and Primates (Springer et al. 2003). However, our previous molecular dating indicates that rodent V1R families expanded after the primate-rodent split (Grus and Zhang 2004), suggesting independent expansions in rodents and primates. (In the future, the independent expansions in primates and rodents could be tested by examining the phylogenetic positions of human pseudogenes when their sequences become available.)

This independence would further imply that some of the aforementioned mammalian orders might not contain expanded V1R families. Even if expansion is characteristic of these five orders, it is also possible that a functional V1R repertoire subsequently shrank after expansion, as in catarrhine primates (Zhang and Webb 2003). These considerations suggest that mouse and rat may be atypical mammals in terms of their vomeronasal receptor genes and vomeronasal-mediated olfactory sensitivities. Of course, independent expansions would also imply great differences in V1R receptors and V1R-mediated olfaction sensitivities. Thus, one should be cautious in applying to other mammals the V1R-related knowledge learned from the model organisms of mouse and rat.

Furthermore, although a large variation in the V1R repertoire is described here, the species we have examined (human, mouse, rat, dog, cow, and opossum) represent only five of ~24 orders of placental and marsupial mammals. A more thorough investigation of the V1R repertoires in other orders will give a better picture of the variation and evolution of V1Rs in mammals.

Interestingly, the small repertoire of V1R receptor genes that we report in the dog may not be entirely unexpected and may be common to carnivores in general. Although rodents have a complex VNO with a thick layer of vomeronasal sensory epithelium (VNSE), where vomeronasal receptors are found, carnivores have a different type of VNO with a much thinner VNSE (Takami 2002; Takigami et al. 2004). Furthermore, the VNO of the ferret *Mustela putorius*, another carnivore, is rudimentary in size and development and does not change with respect to season, which is in contrast to the seasonal variation observed in other ferret organs involved in sexual reproduction and behavior and in rodent VNOs (Weiler et al. 1999). Additionally, Woodley and Baum

(2004) found that the main olfactory system, but not the VNO, is necessary for mate identification in the ferret. The findings in dog and ferret suggest a limited role of the VNO, which may predict a small V1R repertoire in all carnivores. Among artiodactyls, cow is the only species with a fully sequenced genome. However, a few studies have focused on the VNO of the goat, which belongs to the same family as the cow (Takigami et al. 2004; Wakabayashi et al. 2002), and a genomic Southern analysis suggested that the goat V1R repertoire is also significantly smaller than that in rodents (Wakabayashi et al. 2002). However, even between mouse and rat, there is a nearly two-fold difference in the number of functional V1R genes, indicating that ordinal generalizations of V1R repertoire sizes from one species should be drawn cautiously.

As mentioned, the complexity of VNO morphology varies among vertebrates. Takami (2002) classified VNOs into five morphological types from the most complex (A) to the simplest (E). Type A is found in ophidian species such as crotaline, garter, and water snakes. Type B is found in rodents and lagomorphs; though marsupials such as opossum also have VNOs similar to type B. Type C is found in ungulates (e.g., horse, cow, sheep), carnivores (e.g., dog, cat, ferret), prosimian primates, and New World monkeys. Type D is found in amphibians and type E is found in human fetal VNO. Thus, there is a correlation between the morphological complexity of the VNO and the number of intact V1R genes (Fig. 2.4), providing genomic support for the notion that the morphological complexity of the VNO may be used as a proxy for the sophistication of olfactory communications within species, as has been assumed by many anatomists (Giorgi and Rouquier 2002; Wakabayashi et al. 2002).

2.6 ACKNOWLEDGEMENTS

We thank Soochin Cho, Masatoshi Nei, Ondrej Podlaha, David Webb, and three anonymous reviewers for valuable comments. 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, 30430110), Natural Science Foundation of Yunnan, and Chinese Academy of Sciences to Y.P.Z.

Figure 2.1 Variation in the numbers of intact V1R genes (shaded bars) and pseudogenes (open bars) among mouse, rat, human, dog, cow, and opossum, with the phylogenetic relationships among the species shown by the tree. Mouse has 187 functional V1Rs and 168 pseudogenes (Shi et al. 2005; Zhang et al. 2004). Rat has 102 functional V1Rs and at least 50 pseudogenes (Shi et al. 2005, P. Shi unpublished data). Human has 4 V1R ORFs and ~200 pseudogenes (Rodriguez and Mombaerts 2002; Zhang and Webb 2003). Cow has 32 functional V1Rs and 41 pseudogenes (this study). Dog has 8 functional V1Rs and 22 pseudogenes (this study). Opossum has 49 functional V1Rs and 53 pseudogenes (this study). The number of pseudogenes in rat is likely an underestimate as there has been no comprehensive study of rat V1R pseudogenes.

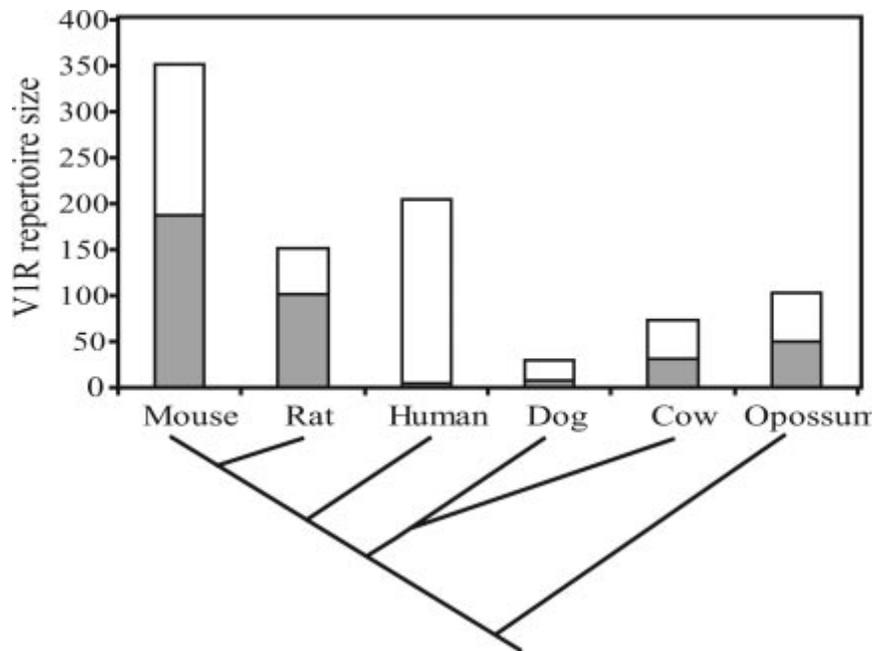


Figure 2.2 Phylogeny of intact V1R genes: 8 in dog, 32 in cow, 4 in human, 102 in rat, and 187 in mouse. Shaded regions group the genes into 18 placental mammalian V1R families previously described (Grus and Zhang 2004; Rodriguez et al. 2002; Shi et al. 2005) or described here (see text), with the family names indicated. Black circles mark family groups that contain more than one family as shown in Table 1. Dog branches are in red, cow branches are in black, human branches are in blue, mouse branches are in purple, and rat branches are in green. Bootstrap percentages supporting the family groups are shown if higher than 50. The tree was reconstructed using the neighbor-joining method with Poisson-corrected protein distances. The arrow points to where the tree is rooted with putative V1Rs of the frog *Xenopus tropicalis* (W. Grus and J. Zhang, unpublished data). The scale bar shows 0.2 amino acid substitutions per site.

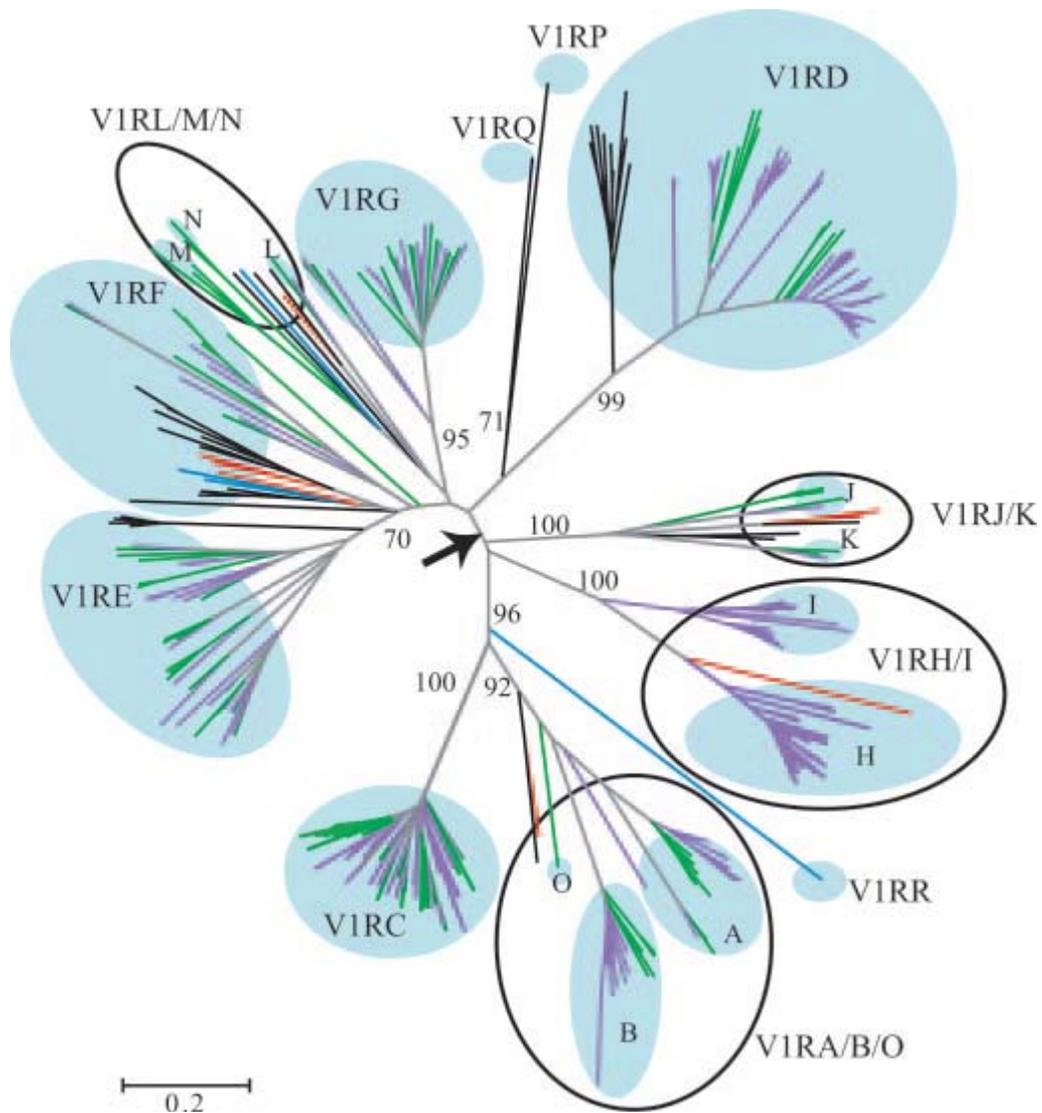


Figure 2.3 Phylogeny of intact V1R genes from the opossum, mouse, rat, dog, and cow. V1R families of placental mammals (V1RA-V1RQ) have been collapsed for better illustration, but all 49 V1Rs of the opossum are shown. Opossum genes are divided into 8 families (oV1RA-oV1RH). The tree was reconstructed using the neighbor-joining method with Poisson-corrected protein distances. Bootstrap percentages are shown if higher than 50. The arrow points to where the tree is rooted with *Xenopus tropicalis* V1Rs. The scale bar shows 0.2 amino acid substitutions per site.

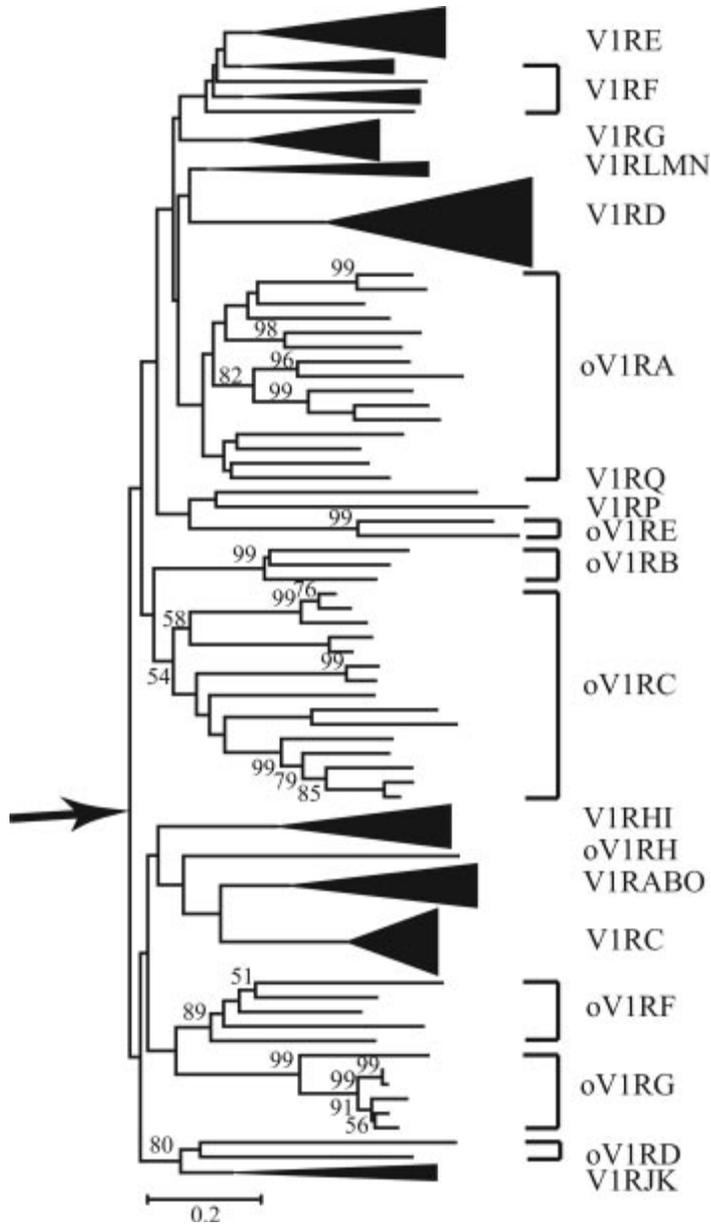


Figure 2.4 Correlation between the morphological complexity of VNO and the number of intact V1R genes in the genome. Morphological complexity follows (Takami 2002).

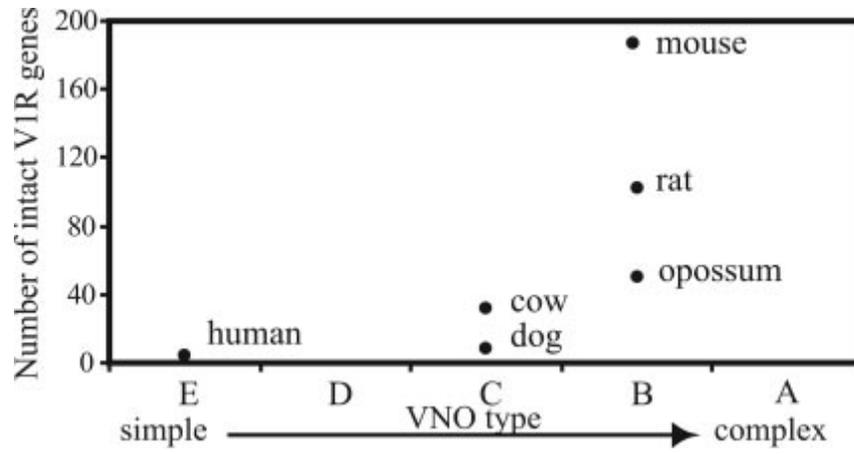


Figure 2.5 Phylogenetic positions of (A) 17 of 22 dog and (B) 35 of 41 cow V1R pseudogenes. Shown here is the unrooted tree of Figure 2.1 with the families collapsed. The phylogenetic position of each pseudogene is determined by making a neighbor-joining tree using the conceptually translated pseudogene sequence with the protein sequences of all intact genes. White families indicate no intact V1R genes in that family group. Gray families indicate a single intact gene in the family group. Black families indicate more than one intact gene in the family group. Pseudogenes are shown in dashed lines. The scale bar shows 0.2 amino acid substitutions per site and is applicable only for branches leading to intact genes.

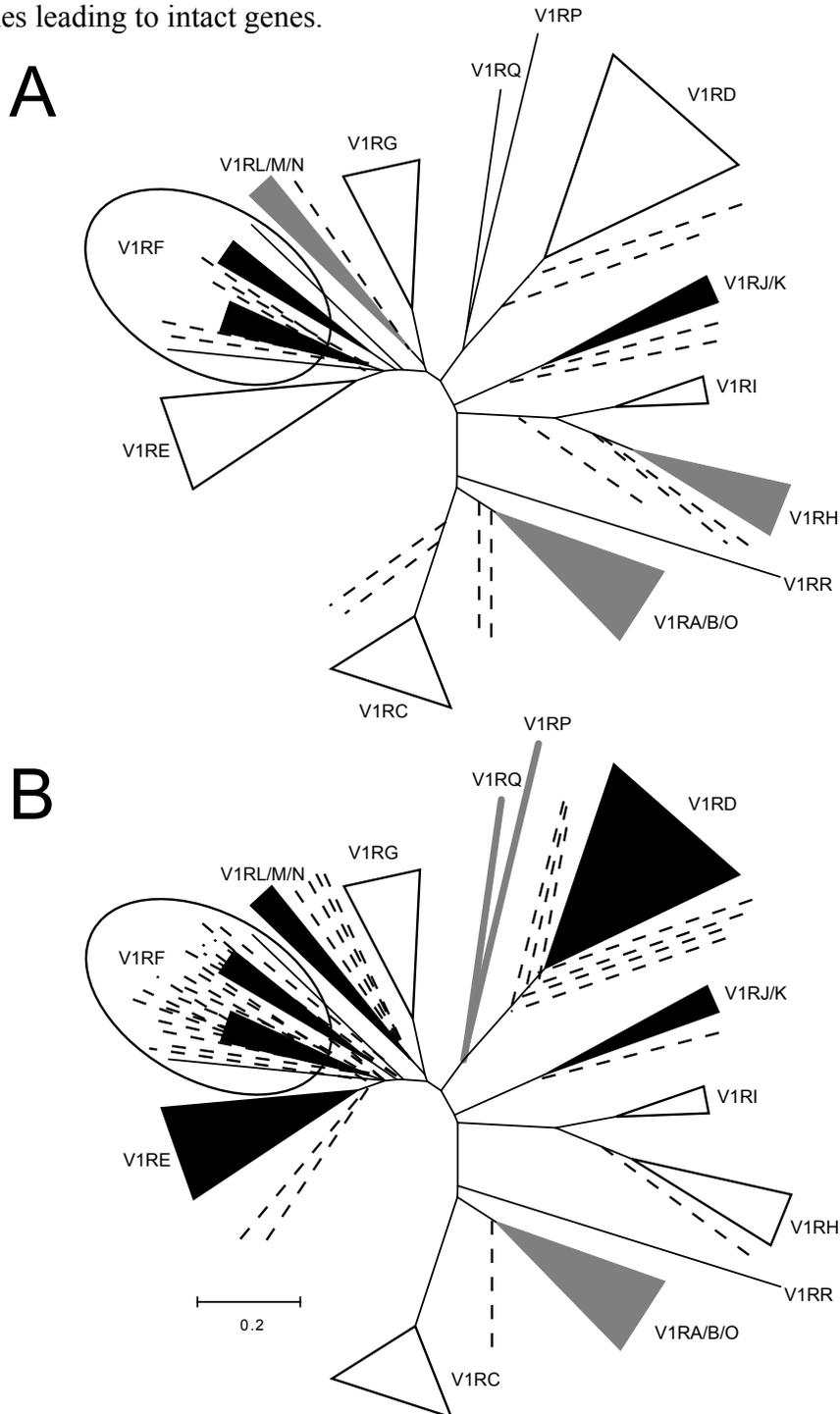


Table 2.1 V1R gene family groups in five placental mammals

Family group	Mouse ¹	Rat ¹	Dog ²	Cow ²	Human ³
A/B/O	19	15	1 (2)	1 (1)	0
C	32	23	0 (2)	0 (0)	0
D	56	8	0 (2)	9 (7)	0
E	16	22	0 (0)	4 (2)	0
F	5	8	3 (5)	10 (19)	2
G	21	13	0 (0)	0 (0)	0
H/I	35	0	1 (3)	0 (1)	0
J/K	2	6	2 (2)	3 (1)	0
L/M/N	1	7	1 (1)	3 (4)	1
P	0	0	0 (0)	1 (0)	0
Q	0	0	0 (0)	1 (0)	0
R	0	0	0 (0)	0 (0)	1
Total	187	102	8 (17)	35 (35)	4

¹ (Grus and Zhang 2004; Shi et al 2005)

² This study; numbers in parentheses are pseudogenes could be classified into family groups. The remaining pseudogenes were too degenerated to determine their family groups.

³ (Rodriguez and Mombaerts 2002; Zhang and Webb 2003)

Table 2.2. Size variation of some gene families among mammalian species.

Gene family ^(ref)	Function	Size variation (ratio)	Smallest size (organism)	Biggest size (organism)
V1R ^{1,this study}	olfactory receptor	23.4	8 (dog) [*]	187 (mouse)
Morpheus ²	nuclear pore complex interacting protein	21	1 (Old World monkeys)	21(chimpanzee)
EAR ^{3,4}	antiparasitic RNases	17	1 (New World monkeys)	17 (ricefield mouse)
Ly49 ⁵	immunity	17	1 (baboon)	17 (rat)
OTEX/PEPP ⁶	reproduction-related	7.5	2 (human)	15 (mouse)
Granzyme ⁷	mast cell chymases	7	4 (human)	28 (rat)
KIR ^{8,9}	immunity	7	2 (mouse)	14 (human, macaque)
OR ^{7,10}	olfactory receptor	3.7	388 (human)	1430 (rat)
Keratin-associated protein ⁷	epithelial cell function	3.3	3 (human)	10 (mouse)
Reverse transcriptase ¹¹	polymerase	2.6	25 (mouse)	65 (human)

Only functional genes are considered. Species with at least one functional gene in the gene family are compared.

Gene families with a size ratio >2 are presented.

^{*}Human V1Rs are likely relics of an ongoing pseudogenization process (Zhang and Webb 2003). Therefore, the dog has the smallest functional repertoire.

¹ (Shi et al 2005)

² (Johnson et al 2001)

³ (Zhang et al 2000)

⁴ (Cho et al 2005)

⁵ (Hao and Nei 2004)

⁶ (LeFranc 2001)

⁷ (Niimura and Nei 2003)

⁸ (Hughes 2002)

⁹ (Hao and Nei 2005)

¹⁰ (Rosenberg et al 1995)

¹¹ (Mouse Genome Sequencing Consortium 2002)

2.7 LITERATURE CITED

- Boschat, C., C. Pelofi, O. Randin, D. Roppolo, C. Luscher, M.C. Broillet, and I. Rodriguez. 2002. Pheromone detection mediated by a V1r vomeronasal receptor. *Nat Neurosci* **5**: 1261-1262.
- Bradley, D.G., D.E. MacHugh, P. Cunningham, and R.T. Loftus. 1996. Mitochondrial diversity and the origins of African and European cattle. *Proc Natl Acad Sci U S A* **93**: 5131-5135.
- Cho, S., J.J. Beintema, and J. Zhang. 2005. The ribonuclease A superfamily of mammals and birds: identifying new members and tracing evolutionary histories. *Genomics*.
- Del Punta, K., T. Leinders-Zufall, I. Rodriguez, D. Jukam, C.J. Wysocki, S. Ogawa, F. Zufall, and P. Mombaerts. 2002. Deficient pheromone responses in mice lacking a cluster of vomeronasal receptor genes. *Nature* **419**: 70-74.
- Dennis, J.C., J.G. Allgier, L.S. Desouza, W.C. Eward, and E.E. Morrison. 2003. Immunohistochemistry of the canine vomeronasal organ. *J Anat* **203**: 329-338.
- Dulac, C. and R. Axel. 1995. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83**: 195-206.
- Dulac, C. and A.T. Torello. 2003. Molecular detection of pheromone signals in mammals: from genes to behaviour. *Nat Rev Neurosci* **4**: 551-562.
- Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783-791.
- Giorgi, D. and S. Rouquier. 2002. Identification of VIR-like putative pheromone receptor sequences in non-human primates. Characterization of VIR pseudogenes in marmoset, a primate species that possesses an intact vomeronasal organ. *Chem Senses* **27**: 529-537.
- Grus, W.E. and J. Zhang. 2004. Rapid turnover and species-specificity of vomeronasal pheromone receptor genes in mice and rats. *Gene* **340**: 303-312.
- Hao, L. and M. Nei. 2004. Genomic organization and evolutionary analysis of Ly49 genes encoding the rodent natural killer cell receptors: rapid evolution by repeated gene duplication. *Immunogenetics* **56**: 343-354.
- Hao, L. and M. Nei. 2005. Rapid expansion of killer cell immunoglobulin-like receptor genes in primates and their coevolution with MHC Class I genes. *Gene*.
- Herrada, G. and C. Dulac. 1997. A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. *Cell* **90**: 763-773.
- Hughes, A.L. 2002. Evolution of the human killer cell inhibitory receptor family. *Mol Phylogenet Evol* **25**: 330-340.
- Hughes, A.L. and M. Nei. 1989. Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. *Mol Biol Evol* **6**: 559-579.
- Johnson, M.E., L. Viggiano, J.A. Bailey, M. Abdul-Rauf, G. Goodwin, M. Rocchi, and E.E. Eichler. 2001. Positive selection of a gene family during the emergence of humans and African apes. *Nature* **413**: 514-519.
- Keverne, E.B. 1999. The vomeronasal organ. *Science* **286**: 716-720.
- Kral, K. 2003. Behavioural-analytical studies of the role of head movements in depth perception in insects, birds and mammals. *Behav Processes* **64**: 1-12.

- Lane, A.J. and D.C. Wathes. 1998. An electronic nose to detect changes in perineal odors associated with estrus in the cow. *J Dairy Sci* **81**: 2145-2150.
- Lane, R.P., T. Cutforth, R. Axel, L. Hood, and B.J. Trask. 2002. Sequence analysis of mouse vomeronasal receptor gene clusters reveals common promoter motifs and a history of recent expansion. *Proc Natl Acad Sci U S A* **99**: 291-296.
- Lane, R.P., J. Young, T. Newman, and B.J. Trask. 2004. Species specificity in rodent pheromone receptor repertoires. *Genome Res* **14**: 603-608.
- Lefranc, M.P. 2001. Nomenclature of the human immunoglobulin heavy (IGH) genes. *Exp Clin Immunogenet* **18**: 100-116.
- Leypold, B.G., C.R. Yu, T. Leinders-Zufall, M.M. Kim, F. Zufall, and R. Axel. 2002. Altered sexual and social behaviors in trp2 mutant mice. *Proc Natl Acad Sci U S A* **99**: 6376-6381.
- Li, W.-H. 1997. *Molecular Evolution*. Sinauer Associates, Sunderland, MA.
- Liman, E.R. and H. Innan. 2003. Relaxed selective pressure on an essential component of pheromone transduction in primate evolution. *Proc Natl Acad Sci U S A* **100**: 3328-3332.
- Matsunami, H. and L.B. Buck. 1997. A multigene family encoding a diverse array of putative pheromone receptors in mammals. *Cell* **90**: 775-784.
- Mombaerts, P. 2004. Genes and ligands for odorant, vomeronasal and taste receptors. *Nat Rev Neurosci* **5**: 263-278.
- Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520-562.
- Mundy, N.I. and S. Cook. 2003. 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**: 1805-1810.
- Murphy, W.J., P.A. Pevzner, and S.J. O'Brien. 2004. Mammalian phylogenomics comes of age. *Trends Genet* **20**: 631-639.
- Nei, M. and A.L. Hughes. 1992. Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. In *Proceedings of the 11th Histocompatibility Workshop and Conference* (eds. K. Tsuji M. Aizawa, and T. Sasazuki), pp. 27-38. Oxford University Press, Oxford.
- Nei, M. and S. Kumar. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- Niimura, Y. and M. Nei. 2003. Evolution of olfactory receptor genes in the human genome. *Proc Natl Acad Sci U S A* **100**: 12235-12240.
- Rat Genome Sequencing Consortium. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**: 493-521.
- Rekwot, P.I., D. Ogwu, E.O. Oyedipe, and V.O. Sekoni. 2001. The role of pheromones and biostimulation in animal reproduction. *Anim Reprod Sci* **65**: 157-170.
- Rodriguez, I., K. Del Punta, A. Rothman, T. Ishii, and P. Mombaerts. 2002. Multiple new and isolated families within the mouse superfamily of V1r vomeronasal receptors. *Nat Neurosci* **5**: 134-140.
- Rodriguez, I. and P. Mombaerts. 2002. Novel human vomeronasal receptor-like genes reveal species-specific families. *Curr Biol* **12**: R409-411.
- Rosenberg, H.F., K.D. Dyer, H.L. Tiffany, and M. Gonzalez. 1995. Rapid evolution of a unique family of primate ribonuclease genes. *Nat Genet* **10**: 219-223.

- Ryba, N.J. and R. Tirindelli. 1997. A new multigene family of putative pheromone receptors. *Neuron* **19**: 371-379.
- Saitou, N. and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406-425.
- Savolainen, P., Y.P. Zhang, J. Luo, J. Lundberg, and T. Leitner. 2002. Genetic evidence for an East Asian origin of domestic dogs. *Science* **298**: 1610-1613.
- Shi, P., J.P. Bielawski, H. Yang, and Y.P. Zhang. 2005. Adaptive diversification of vomeronasal receptor 1 genes in rodents. *J Mol Evol* **60**: 566-576.
- Smit, A.F. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev* **9**: 657-663.
- Springer, M.S., W.J. Murphy, E. Eizirik, and S.J. O'Brien. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc Natl Acad Sci U S A* **100**: 1056-1061.
- Stowers, L., T.E. Holy, M. Meister, C. Dulac, and G. Koentges. 2002. Loss of sex discrimination and male-male aggression in mice deficient for TRP2. *Science* **295**: 1493-1500.
- Su, C. and M. Nei. 1999. Fifty-million-year-old polymorphism at an immunoglobulin variable region gene locus in the rabbit evolutionary lineage. *Proc Natl Acad Sci U S A* **96**: 9710-9715.
- Takami, S. 2002. Recent progress in the neurobiology of the vomeronasal organ. *Microsc Res Tech* **58**: 228-250.
- Takigami, S., Y. Mori, Y. Tanioka, and M. Ichikawa. 2004. Morphological evidence for two types of Mammalian vomeronasal system. *Chem Senses* **29**: 301-310.
- Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876-4882.
- Wakabayashi, Y., Y. Mori, M. Ichikawa, K. Yazaki, and K. Hagino-Yamagishi. 2002. A putative pheromone receptor gene is expressed in two distinct olfactory organs in goats. *Chem Senses* **27**: 207-213.
- Wang, X. and J. Zhang. 2004. Rapid evolution of mammalian X-linked testis-expressed homeobox genes. *Genetics* **167**: 879-888.
- Webb, D.M., L. Cortes-Ortiz, and J. Zhang. 2004. Genetic evidence for the coexistence of pheromone perception and full trichromatic vision in howler monkeys. *Mol Biol Evol* **21**: 697-704.
- Weiler, E., R. Apfelbach, and A.I. Farbman. 1999. The vomeronasal organ of the male ferret. *Chem Senses* **24**: 127-136.
- Wissenbach, U., G. Schroth, S. Philipp, and V. Flockerzi. 1998. Structure and mRNA expression of a bovine trp homologue related to mammalian trp2 transcripts. *FEBS Lett* **429**: 61-66.
- Woodley, S.K. and M.J. Baum. 2004. Differential activation of glomeruli in the ferret's main olfactory bulb by anal scent gland odours from males and females: an early step in mate identification. *Eur J Neurosci* **20**: 1025-1032.
- Zhang, J., K.D. Dyer, and H.F. Rosenberg. 2000. Evolution of the rodent eosinophil-associated RNase gene family by rapid gene sorting and positive selection. *Proc Natl Acad Sci U S A* **97**: 4701-4706.

- Zhang, J. and D.M. Webb. 2003. Evolutionary deterioration of the vomeronasal pheromone transduction pathway in catarrhine primates. *Proc Natl Acad Sci U S A* **100**: 8337-8341.
- Zhang, X., I. Rodriguez, P. Mombaerts, and S. Firestein. 2004. Odorant and vomeronasal receptor genes in two mouse genome assemblies. *Genomics* **83**: 802-811.

CHAPTER 3

LARGEST VERTEBRATE VOMERONASAL TYPE 1 RECEPTOR (V1R) GENE REPERTOIRE IN THE SEMI-AQUATIC PLATYPUS

3.1 ABSTRACT

Vertebrate vomeronasal chemoreception plays important roles in many aspects of an organism's daily life, such as mating, territoriality, and foraging. V1Rs and V2Rs, two large families of G protein-coupled receptors, serve as vomeronasal receptors to bind to various pheromones and odorants. Contrary to previous observations of reduced olfaction in aquatic and semi-aquatic mammals, we here report the surprising finding that the platypus, a semi-aquatic monotreme, has the largest V1R repertoire and nearly largest combined repertoire of V1Rs and V2Rs of all vertebrates surveyed, with 270 intact genes and 579 pseudogenes in the V1R family and 15 intact genes, 55 potentially intact genes, and 57 pseudogenes in the V2R family. Phylogenetic analysis shows a remarkable expansion of the V1R repertoire and a moderate expansion of the V2R repertoire in platypus since the separation of monotremes from placentals and marsupials. Our results challenge the view that olfaction is unimportant to aquatic mammals and call for further study into the role of vomeronasal reception in platypus physiology and behavior.

3.2 INTRODUCTION

Vertebrates use olfaction to locate food, avoid predators, and identify mates, among other activities. Most vertebrates have two olfactory systems: the main olfactory

system (MOS) and the vomeronasal system (VNS). The MOS is traditionally thought to detect environmental odorants while the VNS recognizes intraspecific pheromonal cues, although this distinction has been blurred by recent reports that both systems can perceive both types of signals (Baxi et al. 2006; Restrepo et al. 2004). Nonetheless, a distinct set of chemoreceptors is expressed in each system: odorant receptors (ORs) and trace amine associated receptors (TAARs) in the MOS and vomeronasal type 1 receptors (V1Rs) and type 2 receptors (V2Rs) in the VNS (Liberles and Buck 2006; Mombaerts 2004). These receptors form four evolutionarily unrelated large families of seven-transmembrane G protein-coupled receptors. Our previous study of five orders of placental and marsupial mammals showed that the among-species size variation of the V1R repertoire exceeds that of all other mammalian protein families (Grus et al. 2005). Recently, the genome sequence of the platypus *Ornithorhynchus anatinus*, a semi-aquatic monotreme mammal, became available. It is of significant interest to examine platypus V1Rs for two reasons. First, the platypus represents the final of the three major mammalian groups (placentals, marsupials, and monotremes) whose V1Rs have yet to be examined. Second, olfaction is widely believed to be unimportant to aquatic mammals. For example, morphological components of MOS and VNS are absent or highly degenerated in cetaceans (whales, porpoises, and dolphins) and sirenians (manatees and Dugongs) (Meisami and Bhatnagar 1998; Oelschlager 1992). The platypus data open the door for a genomic test of the (un)importance of olfaction to aquatic mammals, as all previously studied mammalian V1Rs were from terrestrials.

3.3 METHODS

Platypus V1Rs were identified first using TBLASTN searches on the high quality platypus genome sequence (6X coverage) available from the Ensembl Database (http://www.ensembl.org/Ornithorhynchus_anatinus/index.html). Previously described mammalian V1R genes were used as query sequences. Next, BLASTN searches were done on the same platypus genome sequence with the platypus V1R nucleotide sequences from the previous step as query sequences. A receptor having a complete open-reading frame across the middle 13 protein domains (7 transmembrane, 3 extracellular, and 3 intracellular) was considered intact and is most likely functional. V1R pseudogenes were identified by premature stop codons or incomplete sequence across the 13 middle protein domains of the 15 domains. The N-terminal extracellular and C-terminal intracellular domains were not considered in our criterion because they are highly variable in sequence length. A similar two-step procedure was used to identify ORs from platypus and TAAR genes from frog, chicken, platypus, opossum, cow, and dog. Platypus V2Rs were identified using TBLASTN searches on the platypus genome sequences. The more complex structure of the V2Rs requires a few additional steps in gene identification. The computational pipeline for identifying V2Rs from genomic sequence has been outlined in a previous study (Yang et al. 2005). To ensure that the sequences represented independent loci rather than allelic variation, we required that two sequences be at least 2% different at the protein sequence level to be considered as two genes. Protein sequences of the newly identified platypus V1Rs, V2Rs, ORs, and TAARs are available online (<http://mbe.oxfordjournals.org/cgi/content/full/msm157/DC1>) as Supplementary Datasets 1-4, respectively. The V1Rs and V2Rs were aligned by ClustalX (Thompson et

al. 1997) with manual adjustment. Phylogenetic trees were constructed using the neighbor-joining method (Saitou and Nei 1987) with Poisson corrected protein distances and 1000 bootstrap replications (Felsenstein 1985). MEGA3.1 (Kumar et al. 2004) was used for these evolutionary analyses.

3.4 RESULTS AND DISCUSSION

Using TBLASTN searches with known mammalian V1Rs as query sequences, we identified from the platypus genome sequence 270 intact genes and 579 pseudogenes of the V1R family (Table 3.1). To understand the evolution of platypus V1Rs, we reconstructed a gene tree containing all platypus intact V1Rs and 232 representative V1Rs of other vertebrates (Fig. 3.1A). These representatives cover all major lineages in the vertebrate V1R tree (Shi and Zhang 2007). The obtained gene tree shows that platypus V1Rs form three separate platypus-specific clades, two with high bootstrap support, revealing remarkable platypus-specific expansions of three V1R lineages after monotremes diverged from the common ancestor of placentals and marsupials (Fig. 3.1A). Because the draft platypus genome sequence is organized by contig rather than assembled into chromosomes, it is important to ensure that closely related V1Rs are not allelic variants. To be conservative, V1Rs that are >98% identical in amino acid sequence are considered allelic variants (Zhang et al. 2004) and only one is considered in our study. The lack of chromosomal assembly prevents us from detecting V1R genomic clusters that are prevalent in rodents (Grus and Zhang 2004; Lane et al. 2002; Lane et al. 2004; Rodriguez et al. 2002; Zhang et al. 2004). However, many sub-clades in Fig. 3.1A are comprised of V1R genes from a single contig, indicating that there is at least some

chromosomal clustering of closely-related V1R genes in the platypus genome and that tandem gene duplication was a primary mechanism underlying the platypus V1R repertoire expansion. V1R expansions are thought to be correlated to repetitive element density in the V1R genomic regions (Grus et al. 2005; Lane et al. 2002; Lane et al. 2004). Future analyses based on a more accurate genome assembly will likely reveal if any type of repetitive element has a particularly high density in the platypus genomic regions containing V1Rs.

The platypus V1R repertoire has 83 more intact genes than the largest previously identified mammalian V1R repertoire (187 in mouse, Shi et al. 2005). Hence, the mammalian V1R repertoire has an even larger size variation (~34 fold difference between platypus and dog) than was previously reported (~23 fold between mouse and dog, Grus et al. 2005). However, the enormous V1R repertoire in platypus is not entirely unexpected. While previous studies showed the absence or degeneration of MOS and VNS morphological components in some aquatic mammals (Meisami and Bhatnagar 1998), this loss is not observed in the platypus. Instead, VNS morphology in platypus is among the most developed of all mammals, sharing some characteristics with highly complex reptilian VNOs (Wysocki 1979). We previously showed a strong positive correlation between the morphological complexity of the vomeronasal organ (VNO) and the number of intact V1R genes across a diverse array of mammals (Grus et al. 2005). The large V1R repertoire in platypus is consistent with this correlation (Fig. 3.2). Although complete V1R repertoires are known for only 12 vertebrates (Shi and Zhang 2007 and this study), we expect that the platypus repertoire will remain one of the largest

even when additional vertebrates are examined in the future, because of the strong correlation between VNO complexity and V1R repertoire size.

Using an approach similar to Yang et al. (2005), we identified 15 intact genes, 55 potentially intact genes, and 57 pseudogenes of the V2R family from the platypus genome sequence. A potentially intact gene has a partial sequence owing to the incompleteness of the genome sequence, but has an open reading frame in the available sequence. This “partial sequence” problem did not affect the V1Rs in part because they have a single coding exon. Similarly, we reconstructed the phylogeny of the platypus intact V2Rs with 124 representative V2Rs from other vertebrates (Shi and Zhang 2007). Ten of the 15 platypus V2Rs cluster into a single platypus-specific clade, while four other V2Rs form two clusters and one V2R belongs to the V2R2 clade, which has representatives from all vertebrates that have V2Rs (Fig. 3.1B) (Shi and Zhang 2007; Yang et al. 2005; Young and Trask 2007). Thus, at least one V2R lineage experienced a relatively large expansion in platypus after monotremes diverged from placentals and marsupials. In contrast to V1Rs, functional V2R repertoire size is not correlated with VNO complexity (Fig. 3.3). Based on morphology and immunohistochemistry, Takigami et al. (2004) classified mammalian VNOs into two types: segregated and uniform. We recently showed a clear distinction in V2Rs between the two types of VNOs (Shi and Zhang 2007): species with segregated VNOs have functional V2Rs, whereas those with uniform VNOs do not. The platypus VNO has not been characterized in this manner, but our result predicts that it belongs to the segregated type.

Previous studies suggested that V1Rs detect airborne volatiles while V2Rs detect water-soluble molecules (Boschat et al. 2002; Emes et al. 2004; Kimoto et al. 2005;

Leinders-Zufall et al. 2004). Indeed, comparative genomic analysis identified a shift of vomeronasal receptor types from V2Rs to V1Rs during the vertebrate transition from water to land (Shi and Zhang 2007). Unlike most mammals which have a nasal VNO opening with a variable amount of oral input into the nasal cavity, the platypus's complex VNO has an exclusively oral opening, similar to a reptilian VNO. Despite this morphological difference, the platypus VNS likely uses the same VNS signal transduction pathway conserved throughout vertebrate evolution (Grus and Zhang 2006), as other components of the signal transduction pathway (e.g., the *Trpc2* channel) are also found in the platypus genome sequence (data not shown). With the eyes, ears, and nostrils closed while the platypus is underwater (Griffiths 1978), both types of vomeronasal receptors would be exposed to water-soluble molecules. Thus, one may predict that platypus should have experienced a backward evolutionary change, having more V2Rs but fewer V1Rs compared with terrestrial mammals. This, however, is not the case. One potential reason is that the platypus still breaths air and its VNS must still play an important role in recognizing airborne molecules. It is notable that although platypus's V2R repertoire is not large, the total number of intact V1Rs and V2Rs in platypus is 285, nearly highest among all vertebrates surveyed (Table 3.1). If at least half of the potentially intact platypus V2Rs are actually intact, which is quite likely given the current quality of the genome sequence and the length of each partial V2R sequence identified (≥ 234 nucleotides), the total number of platypus V1Rs and V2Rs becomes highest among all vertebrates (Table 3.1).

Given the enormous repertoire size of platypus's VNS receptors, it is interesting to examine whether the opposite is true for its MOS receptors. We raise this possibility

for three related reasons. First, previous studies showed that morphological and/or genetic components of the MOS have been lost or reduced independently in aquatic mammals such as cetaceans, manatees, and seals (Freitag et al. 1998; Meisami and Bhatnagar 1998; Oelschlager 1992). Characterization of the monotreme brain also found that the platypus has a less complex olfactory bulb and olfactory tubercle than terrestrial monotremes (Ashwell 2006a; Ashwell 2006b). Second, in most snakes, particularly marine snakes, the VNS is the dominant chemosensory system and the MOS is very reduced (Evans 2003). In other words, the complex VNS might compensate for a reduced MOS. Indeed, morphological characterization of the platypus brain shows that its accessory olfactory bulb, which is part of the VNS, is larger than its main olfactory bulb, part of the MOS (KW Ashwell, *personal communication*). Third, while terrestrial monotremes rely on their MOS to detect food, platypuses forage for their food underwater with eyes, ears, and nostrils closed (Griffiths 1978). In some aquatic mammals, the loss of olfactory ability is compensated by the enhancement of a different sense, such as audition. Indeed, the monotreme sense of electroreception appears to be more defined in the platypus than in terrestrial monotremes (Pettigrew 1999), and platypuses are thought to use electroreception to locate prey underwater (Scheich et al. 1986).

Applying the same method used for identifying VIRs, we detected 261 intact genes, 94 potentially intact genes, and 221 pseudogenes of the OR family from the platypus genome sequence. The number of intact ORs in platypus is much smaller than that in human, mouse, rat, dog, and frog, although it is greater than that in chicken and fish (reviewed in Niimura and Nei 2006), suggesting that as a mammal, platypus has a

relatively small OR repertoire (Table 3.1). It is notable that the proportion of intact genes in the platypus OR family (45%) is similar to that found in humans, which have a degenerating MOS, and much lower than that (75-80%) in mouse, rat, and dog (Niimura and Nei 2006). While the human OR repertoire is thought to be degenerating due to acquisition of trichromatic vision in catarrhine primates ~23 million years ago (Gilad et al. 2004), fossil evidence suggests that platypuses are a specialized lineage of monotremes that have been semi-aquatic since the Mesozoic (Musser 2003), leaving at least 65 million years for platypus olfactory receptor genes to degenerate with increased aquatic specialization. Furthermore, a fossil platypus from the Miocene shares the extant platypus's reduced morphology of main olfactory brain regions (Macrini et al. 2006). Because the platypus lineage became aquatic a long time ago, many platypus OR pseudogenes may have degenerated beyond detection, leading to an overestimate of the proportion of intact OR genes.

Previous studies of vertebrate ORs describe two classes of ORs: class I is more prevalent in aquatic vertebrates while class II is dominant in terrestrial vertebrates (Freitag et al. 1998; Shi and Zhang 2007). Despite the return to an aquatic life by the platypus, only 11.5% (30 of 261) of the OR genes are the class I aquatic type, which is only slightly higher than that in mouse (11%, Niimura and Nei 2006). This finding may not be unexpected because the platypus nostrils are closed in the water (Griffiths 1978) and the main olfactory epithelium is not exposed to the water. We also examined TAARs, the second class of MOS chemoreceptors (Liberles and Buck 2006), and identified 4 intact genes and 1 pseudogene from the platypus genome. The TAAR repertoire size is also smaller in platypus than in artiodactyls, rodents, and primates

(Table 3.1). Since the platypus VNO opens into the oral cavity, the vomeronasal receptors might also compensate for taste receptors. Preliminary screening of the platypus genome sequence for the T1R sweet and umami receptor family and T2R bitter receptor family reveals a substantively reduced T2R repertoire size in platypus (unpublished data) than in other mammals (Shi and Zhang 2006).

In sum, the investigation of the third major lineage of mammals identified an unexpectedly large repertoire of vomeronasal receptors in the semi-aquatic platypus. This finding challenges the current view that olfaction is unimportant to aquatic mammals and calls for study of the role of vomeronasal reception in platypus physiology and behavior as no study has investigated vomeronasal-mediated behavior in platypus. The surprising diversity of vomeronasal sensitivity across vertebrates provides an invaluable resource for us to learn how nature solves different types of sensory tasks in drastically different environments.

3.5 ACKNOWLEDGEMENTS

We thank Washington University School of Medicine Genome Sequencing Center for making the platypus genome assembly available before publication, Ken Ashwell and Michael Stoddart for valuable communications, and two anonymous reviewers for helpful comments. This work was supported by the National Institutes of Health grant GM80285 to J.Z. W.E.G is supported by NIH training grant T32HG000040.

Figure 3.1. Phylogenetic trees of vertebrate intact (A) V1Rs and (B) V2Rs. The V1R tree includes all 270 platypus receptors and 232 representative receptors from the mouse, rat, dog, cow, opossum, frog, and zebrafish. The V2R tree includes all 15 platypus receptors and 124 representative receptors from the mouse, rat, opossum, frog, and zebrafish, as the dog and cow do not have V2Rs. Pink branches are platypus genes, orange are opossum genes, blue are placental mammal genes, and green are teleost fish and frog genes. Scale bars indicate 0.1 amino acid substitution per site. Bootstrap percentages for clades of platypus receptors are indicated.

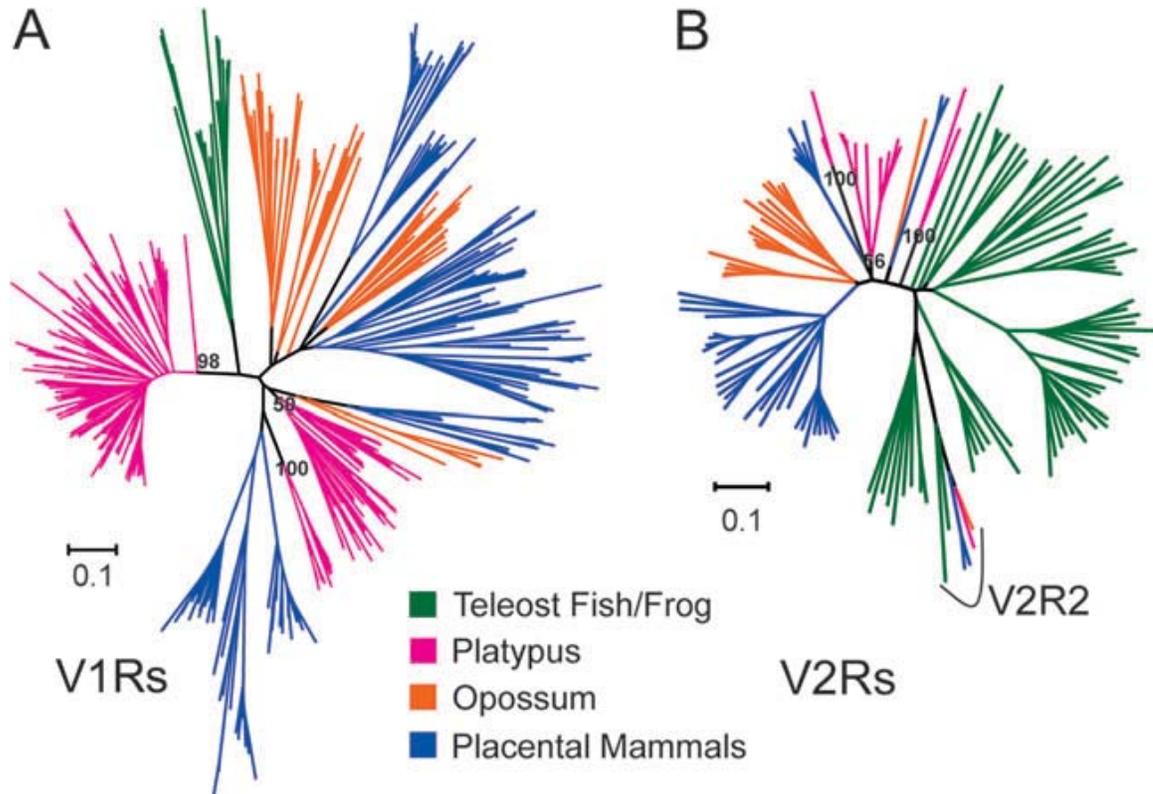


Figure 3.2. Positive correlation of vomeronasal organ complexity with V1R repertoire size. The number of intact V1Rs is plotted for animals of increasing vomeronasal organ complexity (E-A) adapted from morphological categories described in (Takami 2002). The 270 platypus V1Rs are described in this paper while the repertoire sizes for human (5), zebrafish (2), frog (21), dog (8), cow (40), opossum (98), rat (106), and mouse (187) were previously described (Shi and Zhang 2007).

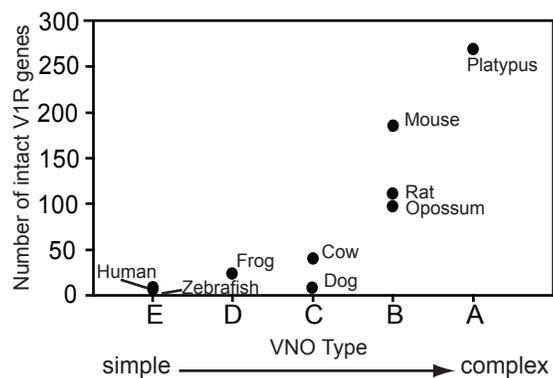


Figure 3.3 No correlation between VNO complexity and V2R repertoire size. The number of intact V2Rs is plotted for animals of increasing vomeronasal organ complexity (E-A) based on morphological categories described by Takami (2002). There is a clear distinction between those mammals with a known segregated VNO (red circles) and those mammals with uniform VNO (blue) based on the classification by Takigami et al. (2004). Species represented by black dots are either not mammals or their VNO structures have not been classified. The 15 platypus V2Rs are described in this paper while the repertoire sizes for zebrafish (44), frog (249), opossum (98), dog (0), cow (0), mouse (121), rat (79), and human (0) were previously described (Shi and Zhang 2007; Young and Trask 2007).

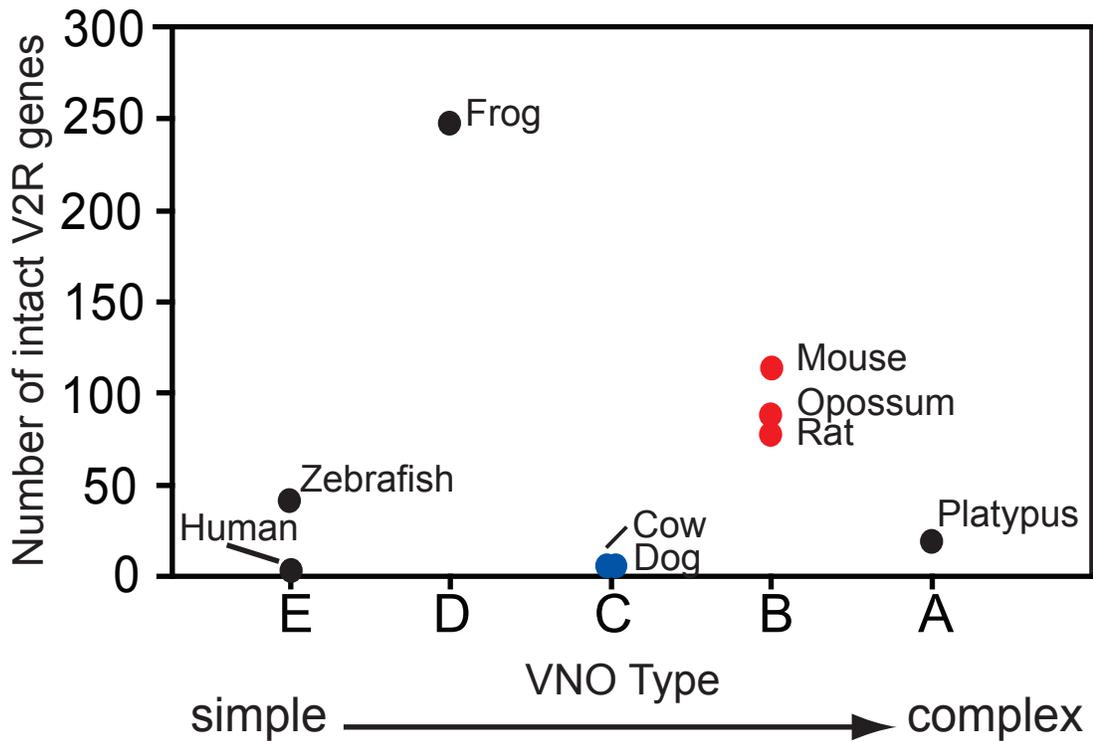


Table 3.1. Sizes of nasal chemosensory receptor gene repertoires in vertebrates.

Species	V1Rs	V2Rs	ORs	TAARs
Human	5 (115) ^a	0 (20) ^c	388 (414) ^d	6 (3) ^h
Mouse	187 (121) ^a	121 (158) ^c	1037 (354) ^d	15 (1) ^h
Rat	106 (66) ^a	79 (142) ^c	1201 (292) ^e	17 (2) ^h
Dog	8 (33) ^a	0 (9) ^c	876 (326) ^f	2 (2) ^b
Cow	40 (45) ^a	0 (16) ^c	970 (1159) ^g	17(9) ^b
Opossum	98 (30) ^a	86 (79) ^c	901 (618) ^f	22 (0) ^b
Platypus	270 (579) ^b	15 (112) ^b	261 (315) ^b	4 (1) ^b
Chicken	0 (0) ^a	0 (0) ^a	82 (476) ^d	3 (0) ^b
Frog	21 (2) ^a	249 (448) ^a	410 (478) ^d	2 (1) ^b
Zebrafish	2 (0) ^a	44 (8) ^a	102 (35) ^d	57 (40) ⁱ

Shown are the numbers of intact genes. The numbers of nonintact genes, including potentially intact genes and pseudogenes, are given in parentheses.

^a Shi and Zhang (2007)

^b This study

^c Young and Trask (2007)

^d Niimura and Nei (2006)

^e Quignon et al (2005)

^f Aloni et al (2006)

^g Niimura and Nei (2007)

^h Lindemann et al (2005)

ⁱ Gloriam et al (2005)

3.6 LITERATURE CITED

- Ashwell, K.W. 2006a. Chemoarchitecture of the monotreme olfactory bulb. *Brain Behav Evol* **67**: 69-84.
- Ashwell, K.W. 2006b. Cyto- and chemoarchitecture of the monotreme olfactory tubercle. *Brain Behav Evol* **67**: 85-102.
- Baxi, K.N., K.M. Dorries, and H.L. Eisthen. 2006. Is the vomeronasal system really specialized for detecting pheromones? *Trends Neurosci* **29**: 1-7.
- Boschat, C., C. Pelofi, O. Randin, D. Roppolo, C. Luscher, M.C. Broillet, and I. Rodriguez. 2002. Pheromone detection mediated by a V1r vomeronasal receptor. *Nat Neurosci* **5**: 1261-1262.
- Emes, R.D., S.A. Beatson, C.P. Ponting, and L. Goodstadt. 2004. Evolution and comparative genomics of odorant- and pheromone-associated genes in rodents. *Genome Res* **14**: 591-602.
- Evans, C. 2003. *Vomeronasal chemoreception in vertebrates: a study of the second nose*. Imperial College Press, London.
- Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783-791.
- Freitag, J., G. Ludwig, I. Andreini, P. Rossler, and H. Breer. 1998. Olfactory receptors in aquatic and terrestrial vertebrates. *J Comp Physiol [A]* **183**: 635-650.
- Gilad, Y., V. Wiebe, M. Przeworski, D. Lancet, and S. Paabo. 2004. Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol* **2**: E5.
- Griffiths, M. 1978. *The Biology of the Monotremes*. Academic Press, New York, NY.
- Grus, W.E., P. Shi, Y.P. Zhang, and J. Zhang. 2005. Dramatic variation of the vomeronasal pheromone receptor gene repertoire among five orders of placental and marsupial mammals. *Proc Natl Acad Sci U S A* **102**: 5767-5772.
- Grus, W.E. and J. Zhang. 2004. Rapid turnover and species-specificity of vomeronasal pheromone receptor genes in mice and rats. *Gene* **340**: 303-312.
- Grus, W.E. and J. Zhang. 2006. Origin and evolution of the vertebrate vomeronasal system viewed through system-specific genes. *Bioessays* **28**: 709-718.
- Kimoto, H., S. Haga, K. Sato, and K. Touhara. 2005. Sex-specific peptides from exocrine glands stimulate mouse vomeronasal sensory neurons. *Nature* **437**: 898-901.
- Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**: 150-163.
- Lane, R.P., T. Cutforth, R. Axel, L. Hood, and B.J. Trask. 2002. Sequence analysis of mouse vomeronasal receptor gene clusters reveals common promoter motifs and a history of recent expansion. *Proc Natl Acad Sci U S A* **99**: 291-296.
- Lane, R.P., J. Young, T. Newman, and B.J. Trask. 2004. Species specificity in rodent pheromone receptor repertoires. *Genome Res* **14**: 603-608.
- Leinders-Zufall, T., P. Brennan, P. Widmayer, P.C. S, A. Maul-Pavicic, M. Jager, X.H. Li, H. Breer, F. Zufall, and T. Boehm. 2004. MHC class I peptides as chemosensory signals in the vomeronasal organ. *Science* **306**: 1033-1037.
- Liberles, S.D. and L.B. Buck. 2006. A second class of chemosensory receptors in the olfactory epithelium. *Nature* **442**: 645-650.

- Macrini, T.E., T. Rowe, and M. Archer. 2006. Description of a cranial endocast from a fossil platypus, *Obdurodon dicksoni* (Monotremata, Ornithorhynchidae), and the relevance of endocranial characters to monotreme monophyly. *J Morphol* **267**: 1000-1015.
- Meisami, E. and K.P. Bhatnagar. 1998. Structure and diversity in mammalian accessory olfactory bulb. *Microsc Res Tech* **43**: 476-499.
- Mombaerts, P. 2004. Genes and ligands for odorant, vomeronasal and taste receptors. *Nat Rev Neurosci* **5**: 263-278.
- Musser, A.M. 2003. Review of the monotreme fossil record and comparison of palaeontological and molecular data. *Comp Biochem Physiol A Mol Integr Physiol* **136**: 927-942.
- Niimura, Y. and M. Nei. 2006. Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates. *J Hum Genet* **51**: 505-517.
- Oelschlager, H. 1992. Development of the olfactory and terminalis systems in whales and dolphins. In *Chemical Signals in Vertebrates* (eds. R. Doty and D. Muller-Schwarze), pp. 141-147. Plenum Press, New York, NY.
- Pettigrew, J.D. 1999. Electroreception in monotremes. *J Exp Biol* **202**: 1447-1454.
- Restrepo, D., J. Arellano, A.M. Oliva, M.L. Schaefer, and W. Lin. 2004. Emerging views on the distinct but related roles of the main and accessory olfactory systems in responsiveness to chemosensory signals in mice. *Horm Behav* **46**: 247-256.
- Rodriguez, I., K. Del Punta, A. Rothman, T. Ishii, and P. Mombaerts. 2002. Multiple new and isolated families within the mouse superfamily of V1r vomeronasal receptors. *Nat Neurosci* **5**: 134-140.
- Saitou, N. and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406-425.
- Scheich, H., G. Langner, C. Tidemann, R.B. Coles, and A. Guppy. 1986. Electroreception and electrolocation in platypus. *Nature* **319**: 401-402.
- Shi, P., J.P. Bielawski, H. Yang, and Y.P. Zhang. 2005. Adaptive diversification of vomeronasal receptor 1 genes in rodents. *J Mol Evol* **60**: 566-576.
- Shi, P. and J. Zhang. 2007. Comparative genomic analysis identifies an evolutionary shift of vomeronasal receptor gene repertoires in the vertebrate transition from water to land. *Genome Res* **17**: 166-174.
- Takami, S. 2002. Recent progress in the neurobiology of the vomeronasal organ. *Microsc Res Tech* **58**: 228-250.
- Takigami, S., Y. Mori, Y. Tanioka, and M. Ichikawa. 2004. Morphological evidence for two types of Mammalian vomeronasal system. *Chem Senses* **29**: 301-310.
- Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876-4882.
- Wysocki, C.J. 1979. Neurobehavioral evidence for the involvement of the vomeronasal system in mammalian reproduction. *Neurosci Biobehav Rev* **3**: 301-341.
- Yang, H., P. Shi, Y.P. Zhang, and J. Zhang. 2005. Composition and evolution of the V2r vomeronasal receptor gene repertoire in mice and rats. *Genomics* **86**: 306-315.
- Young, J.M. and B.J. Trask. 2007. V2R gene families degenerated in primates, dog and cow, but expanded in opossum. *Trends Genet.*

Zhang, X., I. Rodriguez, P. Mombaerts, and S. Firestein. 2004. Odorant and vomeronasal receptor genes in two mouse genome assemblies. *Genomics* **83**: 802-811.

CHAPTER 4

DISTINCT EVOLUTIONARY PATTERNS BETWEEN CHEMORECEPTORS OF TWO VERTEBRATE OLFACTORY SYSTEMS AND THE DIFFERENTIAL TUNING HYPOTHESIS

4.1 ABSTRACT

Most tetrapod vertebrates have two olfactory systems, the main olfactory system (MOS) and the vomeronasal system (VNS). According to the dual olfactory hypothesis, the MOS detects environmental odorants while the VNS recognizes intraspecific pheromonal cues. However, this strict functional distinction has been blurred by recent reports that both systems can perceive both types of signals. Studies of a limited number of receptors suggest that MOS receptors are broadly tuned generalists while VNS receptors are narrowly tuned specialists. However, whether this distinction applies to all MOS and VNS receptors remains unknown. The differential tuning hypothesis predicts that generalist MOS receptors detect an overlapping set of ligands and thus are more likely to be conserved over evolutionary time than specialist VNS receptors, which would evolve in a more lineage-specific manner. Here we test this prediction for all olfactory chemoreceptors by examining the evolutionary patterns of MOS-expressed odorant receptors (ORs) and trace amine associated receptors (TAARs) and VNS-expressed vomeronasal type 1 receptors (V1Rs) and type 2 receptors (V2Rs) in seven tetrapods (mouse, rat, dog, opossum, platypus, chicken, and frog). The phylogenies of V1Rs and V2Rs show abundant lineage-specific gene gains/losses and virtually no one-to-one orthologs between species. Opposite patterns are found for ORs and TAARs. Analysis

of functional data and ligand-binding sites of ORs confirms that paralogous chemoreceptors are more likely than orthologs to have different ligands and that functional divergence between paralogous chemoreceptors is established relatively quickly following gene duplication. Together, these results strongly suggest that the functional profile of the VNS chemoreceptor repertoire evolves much faster than that of the MOS chemoreceptor repertoire and that the differential tuning hypothesis applies to the majority, if not all, of MOS and VNS receptors.

4.2 INTRODUCTION

Most tetrapod vertebrates have two olfactory systems with distinct morphologies and signal transduction pathways: the main olfactory system (MOS) and the vomeronasal system (VNS) (Grus and Zhang 2006). According to the dual olfactory hypothesis (Scalia and Winans 1975), the two systems were thought to have two distinct functions: the MOS detects environmental odorants while the VNS detects intraspecific pheromonal cues. However, exceptions to this distinction began appearing in the literature shortly after the dual olfactory hypothesis was published, and the overlapping functions of the two systems have been the subject of many recent reviews (Restrepo et al. 2004; Baxi, Dorries, and Eisthen 2006; Spehr et al. 2006b; Kelliher 2007). It was found that the MOS sensory neurons can be activated by some pheromones in rabbits and pigs (Hudson and Distel 1983; Dorries, Adkins-Regan, and Halpern 1995) and that the disruption of the MOS signaling pathway affects mouse mating, parenting, and aggressive behaviors, which are widely thought to be mediated by pheromones (Belluscio et al. 1998; Mandiyan, Coats, and Shah 2005). On the other hand, the VNS has been shown to be

important for foraging in some snakes, salamanders, and opossums (Schwenk 1993; Placyk and Graves 2002; Halpern, Daniels, and Zuri 2005). Furthermore, some odorants and pheromones activate neurons in both systems (Sam et al. 2001; Xu et al. 2005; Spehr et al. 2006b; Chamero et al. 2007). These data clearly demonstrate that neither of the two olfactory systems is used exclusively for perceiving one class of chemical cues. Rather, it seems that the two systems are acting in concert to doubly process some chemosignals, albeit through distinct signal transduction pathways (Restrepo et al. 2004; Spehr et al. 2006b).

Despite the fact that the MOS and the VNS have overlapping functions and overlapping ligands, some empirical evidence suggests that VNS receptors may be narrowly tuned to specific ligands while MOS receptors may be broadly tuned to a complex combination of ligands (Leinders-Zufall et al. 2000; Katada et al. 2005). Because each MOS or VNS sensory neuron expresses only one allele of one chemoreceptor gene, the above differential tuning hypothesis applies equally to olfactory neurons and olfactory chemoreceptor proteins. To avoid confusion, we use “receptor” to refer to both neuron and protein and use “chemoreceptor” to refer to protein. It is known that both non-volatile MHC-peptides and the volatile urinary 2-heptanone activate sensory neurons in the MOS and VNS. In the study of VNS activation by MHC-peptides, it was found that important anchoring residues of the peptides were required for signal transduction, that activation was independent of peptide concentration, and that peptides mutated at any site other than the anchoring residues still activated the VNS neurons (Leinders-Zufall et al. 2004). In contrast, changing these key residues had less effect on MOS activation, as some peptides with altered sequences still activated the

MOS neurons if concentration was increased (Spehr et al. 2006a). Similarly, while the VNS response to 2-heptanone was independent of its concentration (Leinders-Zufall et al. 2000), MOS neurons responding to 2-heptanone did so in a concentration dependent manner (Spehr et al. 2006b).

However, it is unknown whether the above observations from a limited number of MOS and VNS responses can be generalized to most or all MOS and VNS receptors. Here we attempt an evolutionary genomic test of the hypothesis of differential tuning between MOS and VNS receptors, by examining the evolutionary patterns of chemoreceptors expressed in the MOS and VNS. Our test is based on the idea that broadly-tuned generalist MOS chemoreceptors are detecting an overlapping set of ligands and should thus be more likely to be conserved over evolutionary time than narrowly-tuned specialist VNS chemoreceptors, which would evolve in a more lineage-specific manner. If an odorant or activating ligand is common to multiple species, the receptor for that odorant is likely to be shared by the multiple species. In contrast, if the odorant yields responses only from a single species, only that species would have to have a receptor that recognizes it. Our test is feasible because the two olfactory systems each express two superfamilies of chemoreceptors. Sensory neurons in the MOS express either odorant receptors (ORs) or trace amine associated receptors (TAARs) (Mombaerts 2004; Liberles and Buck 2006), while those in the VNS express vomeronasal type 1 receptors (V1Rs) or type 2 receptors (V2Rs) (Mombaerts 2004). This expressional distinction is so clear that when the VNS is lost in birds, all V1R and V2R genes were inactivated (Shi and Zhang 2007), whereas OR and TAAR genes are still preserved in the genome (Niimura and Nei 2005; Grus, Shi, and Zhang 2007). The four types of receptors

are all seven-transmembrane G-protein-coupled receptors, but they do not have any significant sequence similarity. Although the four gene superfamilies have been subject to intense evolutionary analysis (Young et al. 2002; Grus and Zhang 2004; Grus et al. 2005; Young et al. 2005; Aloni, Olender, and Lancet 2006; Hoppe et al. 2006; Grus, Shi, and Zhang 2007; Hashiguchi and Nishida 2007; Niimura and Nei 2007; Shi and Zhang 2007; Young and Trask 2007), none of the previous studies has either quantitatively compared the evolutionary patterns of the four superfamilies or compared them in the context of determining their potential functional differences. We show a clear-cut distinction in the evolutionary pattern between MOS and VNS chemoreceptors. Coupled with an analysis of functional data and ligand-binding sites of chemoreceptors, our results provide strong support to the differential tuning hypothesis at the level of entire MOS and VNS chemoreceptor repertoires.

4.3 MATERIALS AND METHODS

4.3.1 Olfactory chemoreceptors from six tetrapods

We compared the V1R, V2R, OR, and TAAR gene repertoires from six tetrapods that have publicly available high-quality genome sequences. These species include frog (*Xenopus tropicalis*), chicken (*Gallus gallus*), platypus (*Ornithorhynchus anatinus*), opossum (*Monodelphis domestica*), dog (*Canis familiaris*), and mouse (*Mus musculus*). We analyzed all previously reported putatively functional olfactory chemoreceptors of these species. The numbers of V1Rs compiled were 187 in mouse (Shi et al. 2005), 8 in dog (Grus et al. 2005; Young et al. 2005), 98 in opossum (Shi and Zhang 2007), 270 in platypus (Grus, Shi, and Zhang 2007), and 21 in frog (Shi and Zhang 2007). Chicken

does not have functional V1Rs. The numbers of V2Rs compiled were 70 in mouse (Yang et al. 2005), 79 in opossum (Shi and Zhang 2007), 249 in frog (Shi and Zhang 2007), and 15 in platypus (Grus, Shi, and Zhang 2007). Chicken and dog do not have functional V2Rs. The numbers of ORs compiled were 1084 in mouse (Young et al. 2002; Zhang and Firestein 2002), 658 in dog (Aloni, Olender, and Lancet 2006), 871 in opossum (Aloni, Olender, and Lancet 2006), 261 in platypus (Grus, Shi, and Zhang 2007), 77 in chicken (Niimura and Nei 2005), and 405 in frog (Niimura and Nei 2005). The numbers of TAARs compiled were 15 in mouse (Lindemann et al. 2005), two in dog (Grus, Shi, and Zhang 2007), 21 in opossum (Grus, Shi, and Zhang 2007), four in platypus (Grus, Shi, and Zhang 2007), three in chicken (Grus, Shi, and Zhang 2007), and three in frog (Grus, Shi, and Zhang 2007).

4.3.2 Rat chemoreceptors

To make the mouse-rat comparison, we obtained 1195 ORs (Rat Genome Sequencing Consortium 2004), 17 TAARs (Lindemann et al. 2005), 106 V1Rs (Shi and Zhang 2007), and 59 V2Rs (Shi and Zhang 2007) from the rat (*Rattus norvegicus*).

4.3.3 Phylogenetic reconstruction

For each of the four olfactory chemosensory receptor superfamilies, protein sequences were aligned by ClustalX (Thompson et al. 1997) with manual adjustment. Phylogenetic trees were reconstructed using the neighbor-joining method (Saitou and Nei 1987) for OR, TAAR, V1R, and V2R superfamilies, respectively, using the six vertebrate species. One thousand bootstrap replications (Felsenstein 1985) were used except in the

OR tree. The comprehensive OR tree had only 400 bootstrap replications while the subtrees constructed for each OR family for better resolution had 1000 bootstrap replications. Trees for the four receptor superfamilies were also reconstructed for the mouse and rat. MEGA (Kumar, Tamura, and Nei 2004) was used for these evolutionary analyses.

4.3.4 Proportion of genes belonging to lineage-specific clades

In the phylogeny for a chemoreceptor superfamily, a lineage-specific clade of n genes must have resulted from at least $n-1$ gene gains/losses since the species diverged from its closest relative in our set of six vertebrates. The total proportion of genes belonging to lineage-specific clades is the sum of these $n-1$ genes for all lineage-specific clades divided by the total number of genes for that chemoreceptor superfamily in that species. A lineage-specific clade is a monophyletic clade of chemoreceptors from a single species. We also used a more stringent definition of lineage-specific clades by the additional requirement that the monophyletic clade should have a bootstrap support of at least 70%. To determine the proportion of MOS chemoreceptors belonging to lineage-specific clades, we summed the $n-1$ genes for a species for both ORs and TAARs and divided it by the total number of ORs and TAARs in that species. A similar approach was used to determine the proportion of VNS chemoreceptors belonging to lineage-specific clades. In the mouse-rat gene tree, a mouse gene and a rat gene are considered to be one-to-one orthologs if they form a monophyletic clade that does not include any other genes. Because orthologous gene pairs are rare for VNS chemoreceptors, to avoid

missing any orthologs, we did not require bootstrap support in defining orthologs in any superfamily to ensure a fair comparison.

4.3.5 Functional comparison of dog ORs

To investigate the relationship between OR sequence divergence and functional divergence, we analyzed a functional dataset of dog ORs (Benbernou et al. 2007). These ORs were originally classified into subfamilies of OR family 6 (Benbernou et al. 2007). However, simple analysis showed that they do not all share the pairwise amino acid sequence identity $\geq 40\%$, which is required for ORs to be classified in the same family (Glusman et al. 2000). To determine their correct classification, we used these ORs to BLAST against the dog ORs in the HORDE database (<http://bioportal.weizmann.ac.il/HORDE/>). Based on our analysis, these 38 dog ORs belong to two class II OR families, OR6 and OR11. We were also able to further classify these ORs into subfamilies based on the criterion that subfamily members share $\geq 60\%$ identity in protein sequence (Glusman et al. 2000). Benbernou et al. (2007) tested the responses of each of these 38 ORs to seven aliphatic aldehydes of C6-C12 at three different concentrations (10^{-6} , 10^{-8} , and 10^{-11} M). We coded each OR's responses by a vector containing the minimal concentration that elicited a response to each aliphatic aldehyde. For no response or response at 10^{-6} , 10^{-8} , or 10^{-11} M, the response was coded as 0, 1, 2, or 3, respectively. For each pair of ORs, we computed the Euclidean distance of their response vectors and compared this functional distance with the proportional difference (*p*-distance) of their amino acid sequences.

4.3.6 Analysis of the OR ligand binding pocket

For each of the 557 pairs of one-to-one mouse-rat orthologous ORs, we randomly picked 10 amino acid sites. For each random set of ten sites, we determined *id*, the number of orthologous pairs that had an identical sequence at these ten sites. We repeated this process 1,000 times for both the entire coding sequence and for only the transmembrane domains to get two distributions for *id*. The simulated values were compared to the actual *id* for the 10 sites in the experimentally determined ligand-binding pocket (Katada et al. 2005). We also conducted a similar analysis for 266 paralogous mouse ORs that belong to the 128 mouse-specific clades in the mouse-rat tree, involving 192 pairwise OR comparisons.

4.4 RESULTS

4.4.1 Distinct phylogenetic patterns between MOS and VNS chemoreceptors

We reconstructed the phylogenies of all putatively functional V1Rs, V2Rs, ORs, and TAARs, respectively, from frog, chicken, platypus, opossum, dog, and mouse (Fig. 4.1). The species were carefully chosen to represent major tetrapod lineages and to avoid overrepresentation of placental mammals, for which many genomes have been sequenced. The difference between the phylogenies of VNS chemoreceptors and MOS chemoreceptors is striking. V1Rs and V2Rs form almost exclusively lineage-specific clades in their trees (Figs. 1A-B), the exception being the V2R2 clade. By contrast, MOS chemoreceptors show a common pattern of multi-taxa clades, although there are also some small lineage-specific clades (Figs. 1C-F).

To quantify the difference in phylogenetic pattern between VNS and MOS chemoreceptors, we calculated the proportion of genes in each chemoreceptor superfamily that arose from lineage-specific gene gains and losses. Using this metric and considering relatively well supported lineage-specific clades (>70% bootstrap), we found a significantly higher proportion of lineage-specific chemoreceptors in the VNS than in the MOS for frog ($\chi^2 = 11.89$, $P < 0.001$), platypus ($\chi^2 = 127.99$, $P < 10^{-28}$), opossum ($\chi^2 = 120.76$, $P < 10^{-27}$), and mouse ($\chi^2 = 201.28$, $P < 10^{-44}$) (Fig. 4.2A). The pattern is less prominent for the frog than for the mammals, probably because the frog lineage is so long (Fig. 4.2A) that it has less power to differentiate frog-specific events from other lineage-specific events. Similar results were obtained when all lineage-specific clades were considered regardless of the bootstrap support (Fig. 4.5).

In the above analyses, we used six distantly related species to represent major tetrapod lineages. To examine if the distinct evolutionary patterns of VNS and MOS chemoreceptors observed at this large evolutionary distance also occurs between closely related taxa, we compared the mouse and rat, two species that diverged approximately 18 million years ago (Murphy, Pevzner, and O'Brien 2004). We constructed new trees using mouse and rat chemoreceptors (Fig. 4.6) and calculated the same metric for a comparison of the four chemoreceptor superfamilies in the two species. Again, we found a higher proportion of genes from lineage-specific gains and losses in the VNS chemoreceptors than in the MOS chemoreceptors for both mouse ($\chi^2 = 401.82$, $P < 10^{-88}$) and rat ($\chi^2 = 52.48$, $P < 10^{-12}$) (Fig. 4.2B). We also determined the number (m) of potentially orthologous gene pairs between the two species in each chemoreceptor superfamily by counting the number of monophyletic clades consisting of one mouse gene and one rat

gene (Appendix A1). We then calculated the fraction (f) of genes in each chemoreceptor superfamily that have one-to-one mouse-rat orthologs, using $f = m/\min(x, y)$, where x and y are the numbers of mouse and rat genes in the superfamily, respectively. The f values for V1Rs (18/106=0.17) and V2Rs (4/59=0.07) are significantly lower than those for ORs (557/1084=0.51) and TAARs (7/15=0.47) ($P < 0.015$ in each of the four comparisons, Fisher's exact test).

Although about half of mouse ORs and TAARs have potential orthologs in rat, these genes with orthologs are not distributed evenly among different ORs and TAARs. Tetrapod ORs have been classified into two classes: fish-like Class I ORs and terrestrial Class II ORs (Freitag et al. 1995; Mombaerts 2004). The f value is 76/123=0.62 for Class I ORs, significantly greater than that (481/961=0.50) for Class II ORs ($P < 0.01$, Fisher's exact test). Mammalian TAARs have also been classified into three families: TAAR1-4, TAAR5, and TAAR6-9 (Lindemann et al. 2005). All mouse and rat TAARs in families TAAR1-4 and TAAR5 are part of potentially orthologous gene pairs. In contrast, only four of the 22 mouse and rat TAARs in family TAAR6-9 are part of potentially orthologous gene pairs. Interestingly, family TAAR6-9 also has had independent expansion in the opossum lineage (Grus, Shi, and Zhang 2007; Hashiguchi and Nishida 2007).

4.4.2 Functional divergence after chemoreceptor gene duplication

It is generally believed that paralogous proteins are much more likely than one-to-one orthologs to have divergent functions (Zhang 2003). This hypothesis can be verified for the chemoreceptors. Functional evidence based on site-directed mutagenesis studies

and computational predictions suggests that the ligand-binding domain of ORs lies in a hydrophobic binding pocket created in the transmembrane domains (Baldwin 1994; Man, Gilad, and Lancet 2004; Zhang et al. 2004; Katada et al. 2005). Specifically, ten residues in transmembranes 3, 5, and 6 in mouse OR 73 (in the OR5D subfamily) were identified to constitute its ligand binding pocket (Katada et al. 2005). Site-directed mutagenesis at these residues altered the ligand binding profile of mouse OR 73 to eugenol and related ligands, suggesting that these residues play a role in OR recognition of ligands. We found that, between mouse-rat orthologous ORs, the 10 ligand-binding sites are more conserved than randomly chosen 10 sites from the entire protein ($P = 0.003$) or from the transmembrane domains ($P = 0.012$) (Fig. 4.3). For the ligand-binding sites, 78.1% (435/557) of orthologous mouse-rat OR pairs have identical sequences, while this fraction is on average 55.3% for 10 randomly chosen sites from the entire protein or 60.3% for 10 randomly chosen sites from the transmembrane domains. In contrast, when we examined ORs that belong to the mouse-specific clades in the mouse-rat OR tree (Fig. 4.6C), the 10 ligand-binding sites are more variable than randomly chosen sites from the entire protein ($P = 0.098$) or the transmembrane domains ($P = 0.128$). Thus, consistent with the expectation, this comparison suggests that one-to-one orthologous ORs tend to recognize the same ligands, whereas paralogous ORs tend to recognize different ligands.

The above results, however, do not tell how quickly a newly duplicate OR establishes its function. Here we compare the relationship between OR sequence divergence (proxy for time) and functional divergence using a functional dataset of 38 paralogous ORs from the dog (Benbernou et al. 2007). Each of these 38 ORs was previously examined for response to seven aliphatic aldehydes of C6-C12 at three

different concentrations (Benbernou et al. 2007). For dog ORs belonging to the same subfamily, OR functional divergence was positively correlated with protein sequence divergence ($R^2 = 0.155$; $P < 0.01$; Fig. 4.4). In other words, more divergent ORs responded more differently to ligands. However, for ORs belonging to different subfamilies or different families, no correlation was observed between sequence divergence and functional divergence ($R^2 = 0.000132$; $P = 0.77$; Fig. 4.4). These observations suggest that functional changes may occur only in newly duplicated OR genes and that once the function is established in an OR, it is no longer altered.

4.5 Discussion

The two vertebrate olfactory systems, the VNS and the MOS, overlap in some of their activating ligands. But our analysis showed that they differ greatly in the evolutionary patterns of their chemoreceptors. While the phylogenies of VNS chemoreceptors exhibit almost exclusively lineage-specific clades, those of MOS chemoreceptors show both multi-species clades and lineage-specific clades. One potential caveat is that each chemoreceptor gene superfamily studied here form one to multiple gene clusters in chromosomes and different chromosomal regions may have different intrinsic rates of gene duplication. However, it seems unlikely that by chance the two VNS chemoreceptor superfamilies are both in high duplication regions while the two MOS chemoreceptor gene superfamilies are both in low duplication regions. This possibility becomes even lower when we consider that V1R, V2R, and OR genes are actually scattered in multiple chromosomes (Young et al. 2002; Zhang and Firestein 2002; Grus and Zhang 2004; Zhang et al. 2004; Yang et al. 2005). Thus, the contrast

between the evolutionary patterns of MOS and VNS receptors reflects a difference at the selection level rather than the mutation level.

Although all four chemoreceptor families we investigated are present in teleost genomes (Alioto and Ngai 2005; Niimura and Nei 2005; Alioto and Ngai 2006; Hashiguchi and Nishida 2007; Pfister et al. 2007; Shi and Zhang 2007), we focused our comparative analysis on tetrapod taxa. In teleost fish, both the VNS and MOS are expressed in sensory neurons in the olfactory epithelium because teleost fish do not have a separate VNO. It has already been shown that there were shifts in olfactory chemoreceptor prevalence following the vertebrate transition from aquatic habitats to terrestrial habitats with Class I ORs and V2Rs dominating primary aquatic vertebrates and Class II ORs and V1Rs dominating in terrestrial vertebrates (Shi and Zhang 2007). Functional changes of the chemoreceptors likely accompanied this evolutionary shift. For example, based on the comparative sequence analysis, teleost V2Rs are thought to be amino acid receptors while mammalian V2Rs do not contain the conserved residues necessary for amino acid binding (Alioto and Ngai 2006). Additionally, In contrast to patterns observed in mammalian V1Rs, teleost V1Rs are highly conserved among distantly related taxa (Saraiva and Korsching 2007), suggesting a different role for these teleost chemoreceptors. Furthermore, the difference in environment between aquatic and terrestrial vertebrates has altered the nature of the ligands, such that classes of common teleost odorants, such as bile acids, amino acids, steroids, and prostaglandins, are not all common classes of tetrapod odorants.

The basis of our analysis is in recognizing lineage-specific clades. We defined a lineage-specific clade as a monophyletic clade of multiple receptors, all from one species.

Such clades were formed by (1) gene duplication that occurred after the evolutionary separations of the species studied here, (2) loss of ancestral genes in specific lineages, or (3) a combination of the above two processes. The majority of our focus lies in the first case, although all three are likely to occur. The striking difference in the evolutionary pattern between VNS and MOS chemoreceptors indicates substantively faster changes of the VNS receptor repertoire than the MOS receptor repertoire during evolution. The VNS chemoreceptor superfamilies acquired and lost genes with such a high rate that there are no one-to-one orthologous chemoreceptors among the initial six tetrapod species examined, except for the V2R2 clade, which has a different evolutionary origin (Yang et al. 2005; Shi and Zhang 2007), different expression pattern and transport mechanism (Martini et al. 2001; Silvotti, Giannini, and Tirindelli 2005), and possibly different function (Young and Trask 2007) from other V2Rs. Even between the closely related mouse and rat, less than 11% of VNS chemoreceptors have one-to-one orthologs, compared to over 48% of MOS chemoreceptors. These results are in stark contrast to other groups of GPCRs, which all have at least 84% one-to-one orthologs between mouse and rat (Gloriam, Fredriksson, and Schioth 2007). Our analysis of OR functions and ligand-binding sites showed that paralogous chemoreceptors are much more likely than one-to-one orthologs to have divergent functions and that the functional divergence tends to be established shortly after gene duplication. Taken together, our results suggest that the functional profile of the VNS receptor repertoire evolves much faster than that of the MOS receptor repertoire, which is consistent with the prediction of the differential tuning hypothesis that broadly-tuned generalist MOS receptors detect an overlapping set of

ligands and thus are more likely to be conserved over evolutionary time than narrowly-tuned specialist VNS receptors, which would evolve in a more lineage-specific manner.

There are additional lines of evidence for the differential tuning hypothesis. In the study that identified the 10 ligand-binding pocket residues of mouse OR 73, Katada et al (2005) aimed to understand the molecular mechanism allowing for the identification of tens of thousands of potential ligands by ~1000 ORs. The complex combinatorial mechanism of OR-ligand binding by which each OR can bind multiple ligands and each ligand can be bound by multiple ORs (Malnic et al. 1999) is much different from the highly specific receptor-ligand pairing for other GPCRs (Katada et al. 2005). These authors showed that the majority of ligand-binding residues for ORs are nonpolar, and the majority of interactions between ORs and ligands are weak hydrophobic or van der Waals interactions (Katada et al. 2005). In contrast, they note that the β 2-adrenergic receptor has polar and charged residues in its ligand binding pocket, allowing for stronger ionic interactions between ligand and receptor and higher ligand affinities (Wieland et al. 1996; Katada et al. 2005). Although the ligand binding pocket has not been identified for either class of VNS chemoreceptor, we found that transmembranes 3, 5, and 6 have a significantly higher number of polar residues (ORs mean=34.2%, V1Rs mean=37.3%; Fisher's exact test, $P < 10^{-28}$) and charged residues (ORs means=6.6%, V1Rs mean=7.2%; Fisher's exact test, $P < 10^{-4}$) in V1Rs (from mouse, dog, opossum, platypus, and frog) than in ORs (from the same species), suggesting that there is potential for stronger and more specific ligand-receptor interactions in V1Rs than in ORs, consistent with the differential tuning hypothesis.

A number of functional studies of ORs are consistent with the hypothesis of differential tuning in VNS and MOS chemoreceptors. For instance, the same ligands can activate distantly related ORs (Malnic et al. 1999; Sanz et al. 2005; Benbernou et al. 2007; Stary et al. 2007). In contrast, only a few VNS receptors are activated by 2-heptanone and ESP1 (Leinders-Zufall et al. 2000; Kimoto et al. 2005). Additionally, all VNS neurons that respond to the same ligand have the same response profile, not responding to any of the other V1R ligands (Leinders-Zufall et al. 2000). In contrast, ORs that respond to octanal had a wide-range of response profiles to related odorants (Araneda et al. 2004; Benbernou et al. 2007). Furthermore, concordant birth-and-death evolutionary patterns between the V2R superfamily and the two V2R ligand families strongly suggest high ligand specificity of VNS chemoreceptors (Chamero et al. 2007; Shi and Zhang 2007). We conclude that multiple lines of evidence, particularly the distinct evolutionary patterns of MOS and VNS chemoreceptors, strongly support differential tuning between MOS and VNS receptors at the level of entire receptor repertoires.

4.6 ACKNOWLEDGMENTS

We thank Meg Bakewell, Peng Shi, and Ben-Yang Liao for valuable comments. This work was supported by National Institutes of Health (NIH) grant R01GM80285 to J.Z. W.E.G was supported by NIH training grant T32HG000040.

Figure 4.1 Unrooted phylogenetic trees of all putatively functional frog, chicken, platypus, opossum, dog, and mouse (A) V1Rs, (B) V2Rs, (C) ORs, (D) Class I ORs, (E) Class II ORs from families 5 and 8 as defined by the HORDE database, and (F) TAARs. Because there are too many Class II ORs to show clearly, families 5 and 8 are randomly chosen to illustrate the phylogenetic pattern of Class II ORs. Other families show a similar pattern. The trees were reconstructed using the neighbor-joining method with Poisson-corrected protein distances. The scale bars show 0.1 amino acid substitutions per site. The phylogenetic patterns show that vomeronasal system (VNS) chemoreceptors tend to form species-specific clades, which are rarely found among MOS chemoreceptors.

Figure 4.1

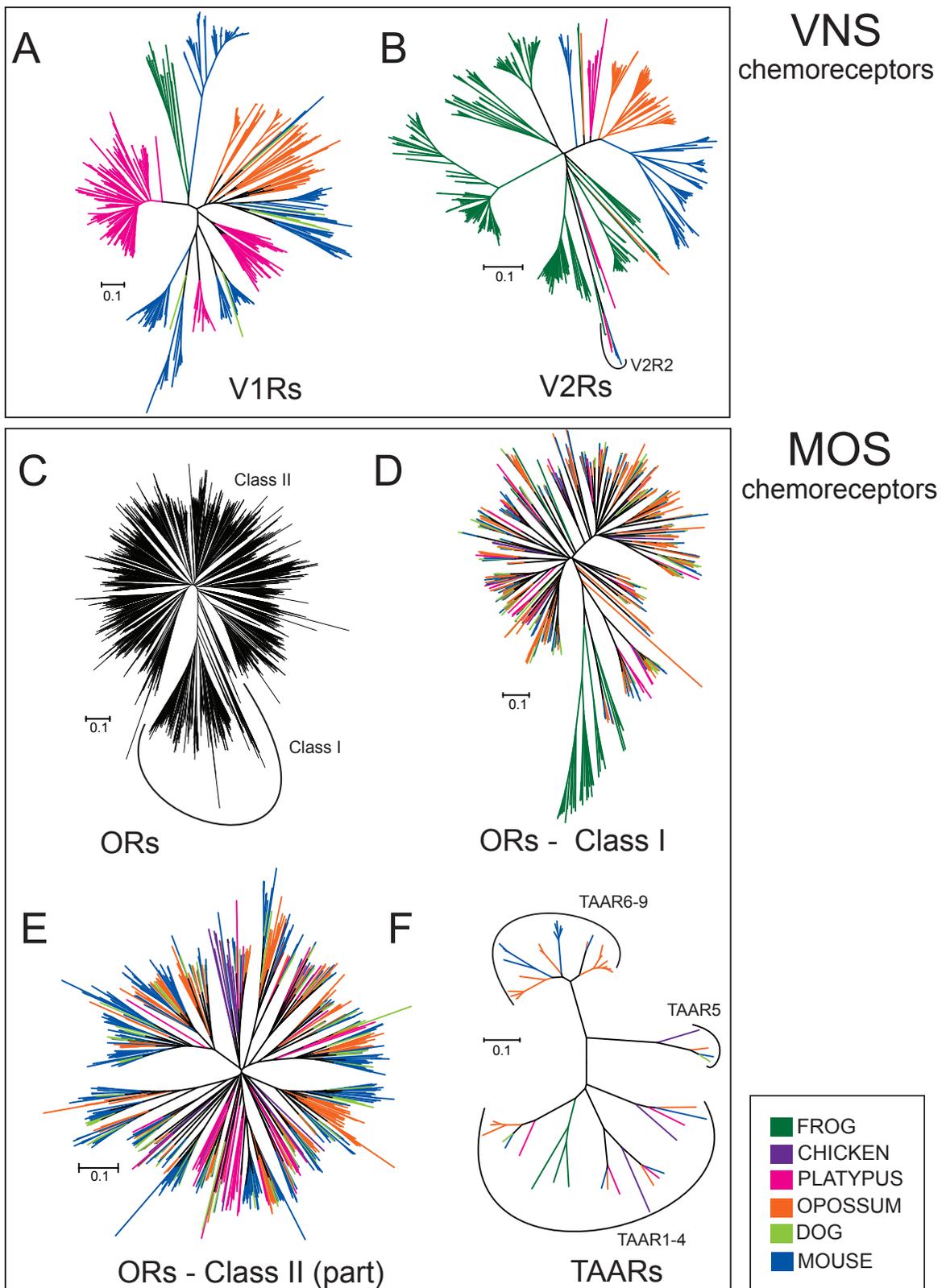


Figure 4.2 Proportion of genes belonging to lineage-specific clades \pm one standard error in MOS receptors (white bars) and VNS receptors (gray bars) in (A) the distantly related species frog, platypus, opossum, and mouse and (B) the closely related species mouse and rat. Statistical significant differences are indicated by * for $P < 10^{-3}$, ** for $P < 10^{-10}$, and *** for $P < 10^{-20}$. Data from ORs and TAARs are combined for MOS chemoreceptors, and data from V1Rs and V2Rs are combined for VNS chemoreceptors. Bolded lines on the trees indicate the lineages for which lineage-specific clades are defined. Dotted lines lead to taxa that are included in phylogenetic analysis (Fig. 4.1) but are not presented here because they lack functional members in one or more chemoreceptor families. Scale bars indicate evolutionary times (million years ago, MYA) (Murphy et al. 2004). Mouse in (A) represents the mouse lineage since its separation from the dog lineage. Mouse in (B) represents the mouse lineage since its separation from the rat lineage.

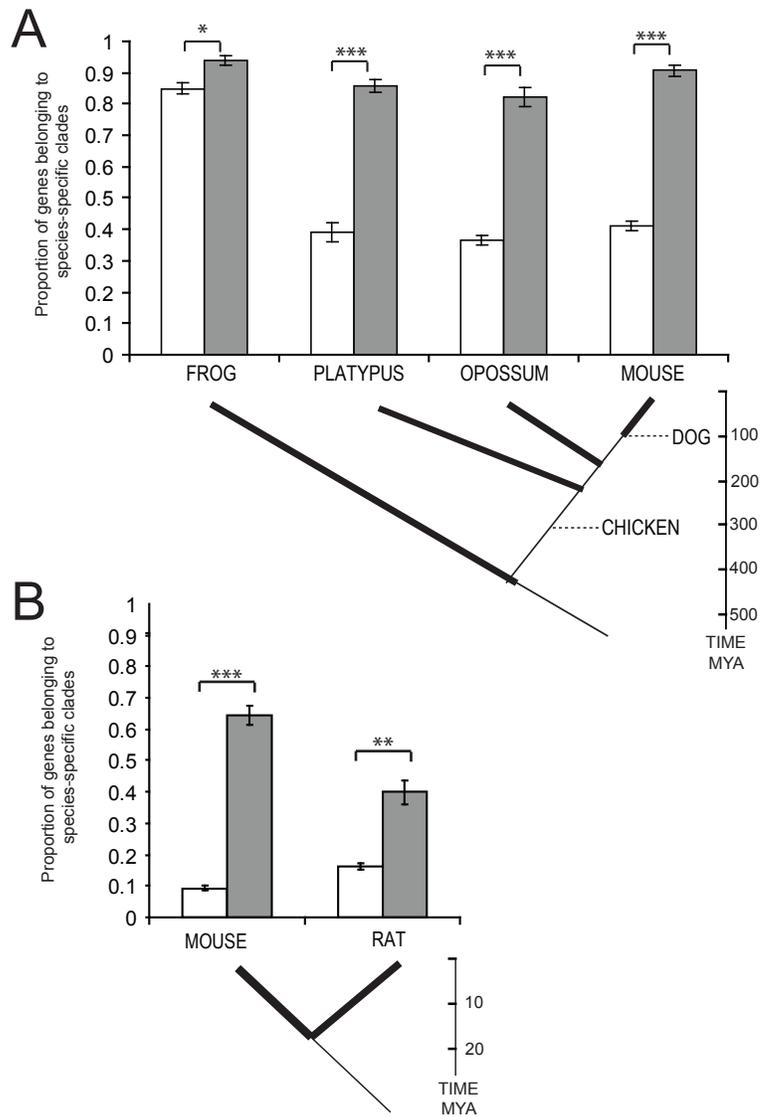


Figure 4.3 Ligand-binding sites are conserved in mouse-rat orthologous ORs, but not in lineage-specific paralogous mouse ORs. (A) Frequency distribution of the number of orthologous mouse-rat ORs with identical sequence at 10 randomly chosen sites, derived from 1000 random samples of 10 sites. A total of 557 mouse-rat orthologous OR pairs are compared. The number of orthologous ORs with identical sequences at the 10 ligand-binding sites (435) is significantly higher than that at 10 randomly chosen sites from the entire protein (mean=308; $P = 0.003$, one-tail test) or from the transmembrane domains (mean=336; $P = 0.012$, one-tail test). (B) Frequency distribution of the number of paralogous mouse ORs with identical sequence at 10 randomly chosen sites, derived from 1000 random samples of 10 sites. A total of 192 pairwise comparisons involving 266 mouse ORs in 128 mouse-specific clades of the mouse-rat OR tree are conducted. The number of paralogous ORs with identical sequences at the 10 ligand-binding sites (96) is smaller than that at 10 randomly chosen sites from the entire protein (mean=114; $P = 0.098$, one-tail test) or from the transmembrane domains (mean=114; $P = 0.128$, one-tail test).

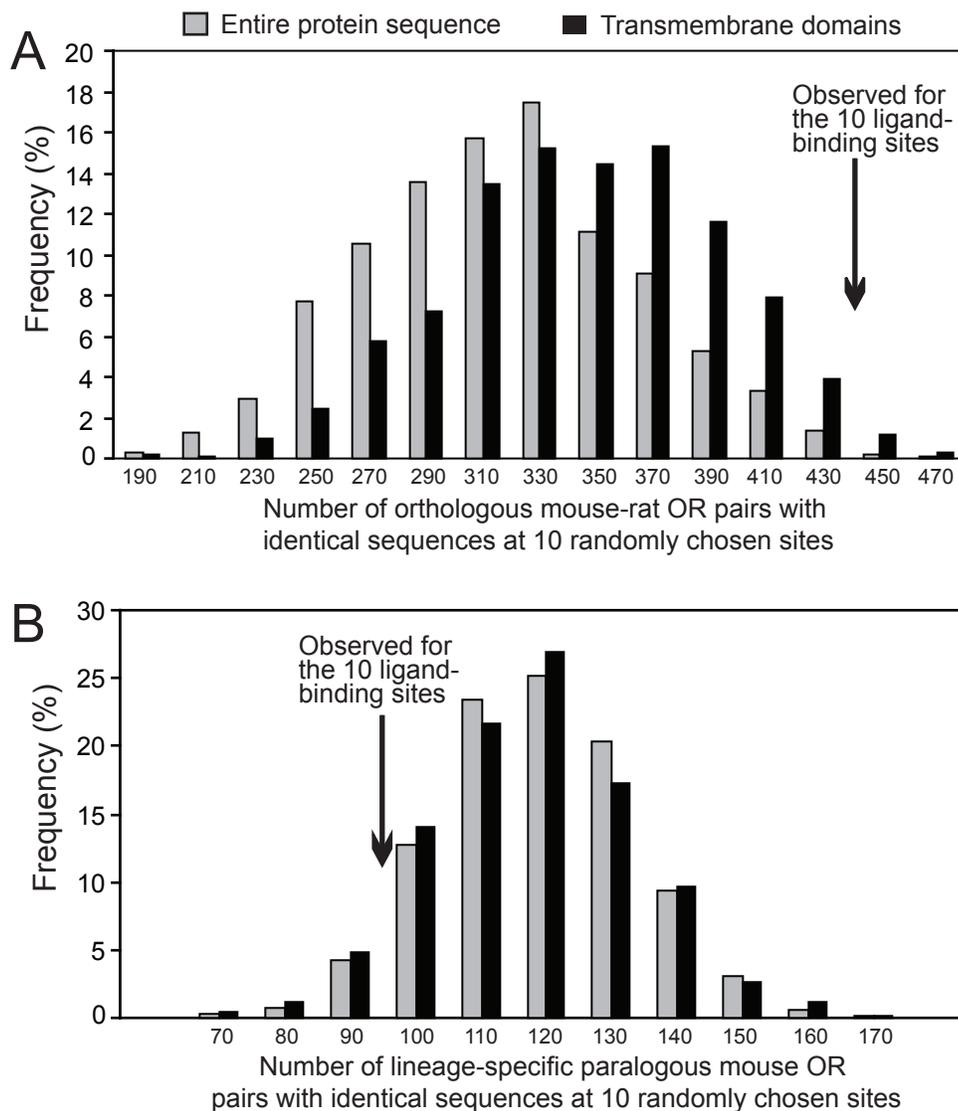


Figure 4.4 Pairwise amino acid sequence divergence and OR response-profile distance are significantly correlated between closely related dog OR paralogs (pluses), but not between more distantly related OR paralogs (circles). OR response data taken from (Benbernou et al. 2007).

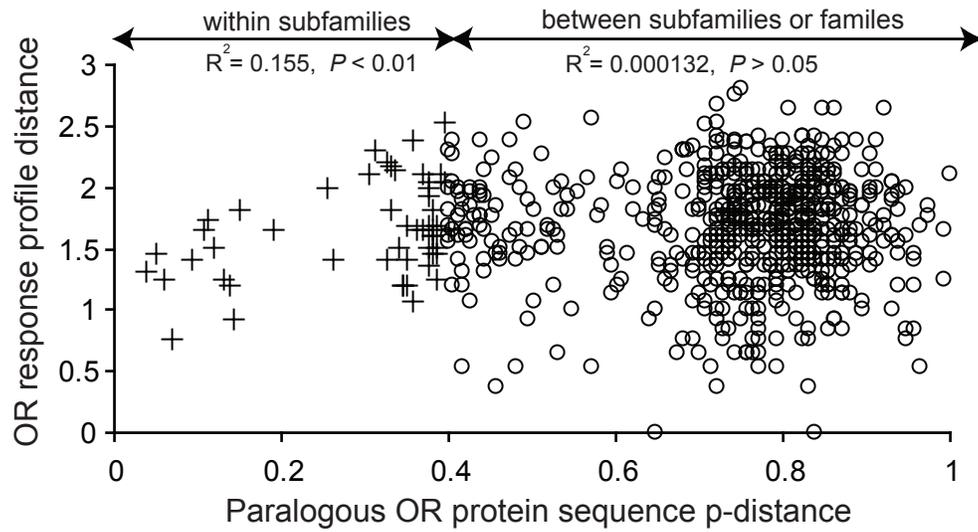


Figure 4.5 Proportion of genes belonging to lineage-specific clades \pm one standard error in MOS chemoreceptors (white bars) and VNS chemoreceptors (gray bars) in (A) the distantly related species frog, platypus, opossum, and mouse and (B) the closely related species mouse and rat. Statistical significant differences are indicated by * for $P < 10^{-3}$, ** for $P < 10^{-10}$, and * for $P < 10^{-20}$. The significance level is 0.06 for the frog. Data from ORs and TAARs are combined for MOS chemoreceptors, and data from V1Rs and V2Rs are combined for VNS chemoreceptors. Bolded lines on the trees indicate the lineages for which lineage-specific clades are defined. Dotted lines lead to taxa that are included in phylogenetic analysis (Fig. 4.1) but are not presented here because they lack functional members in one or more chemoreceptor families. Scale bars indicate evolutionary times (million years ago, MYA) (Murphy et al. 2004). Mouse in (A) represents the mouse lineage since its separation from the dog lineage. Mouse in (B) represents the mouse lineage since its separation from the rat lineage. These results differ slightly from Fig. 4.2 in that all species-specific clades, regardless of bootstrap support, are considered.**

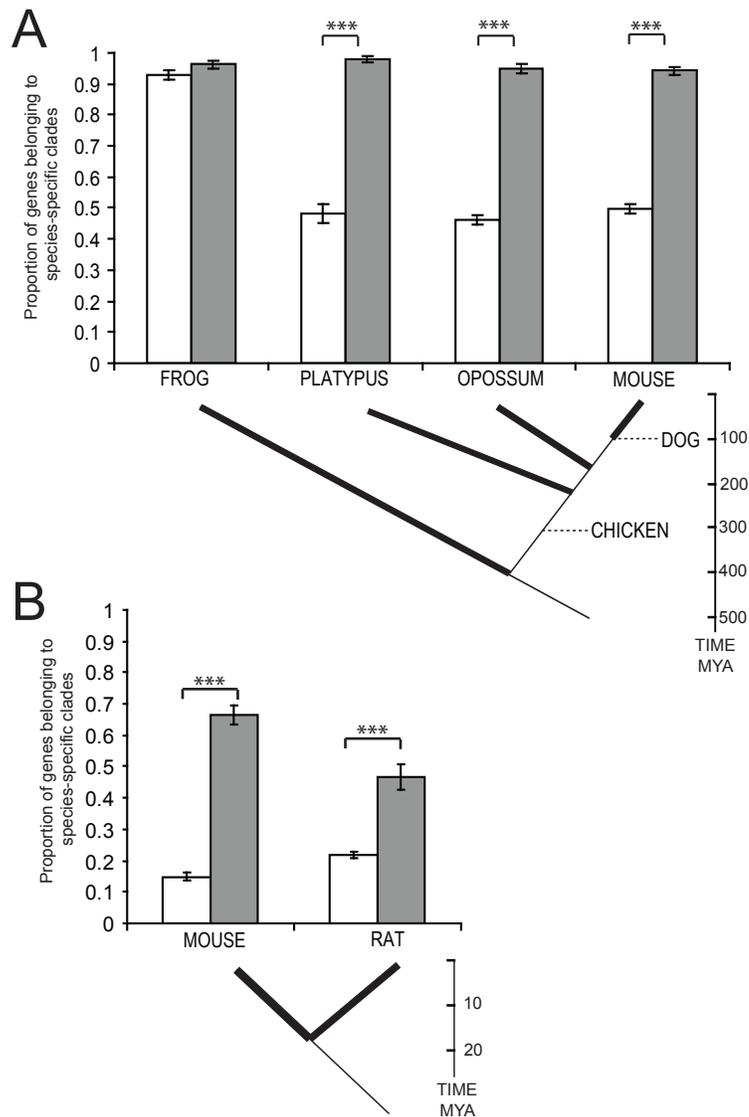
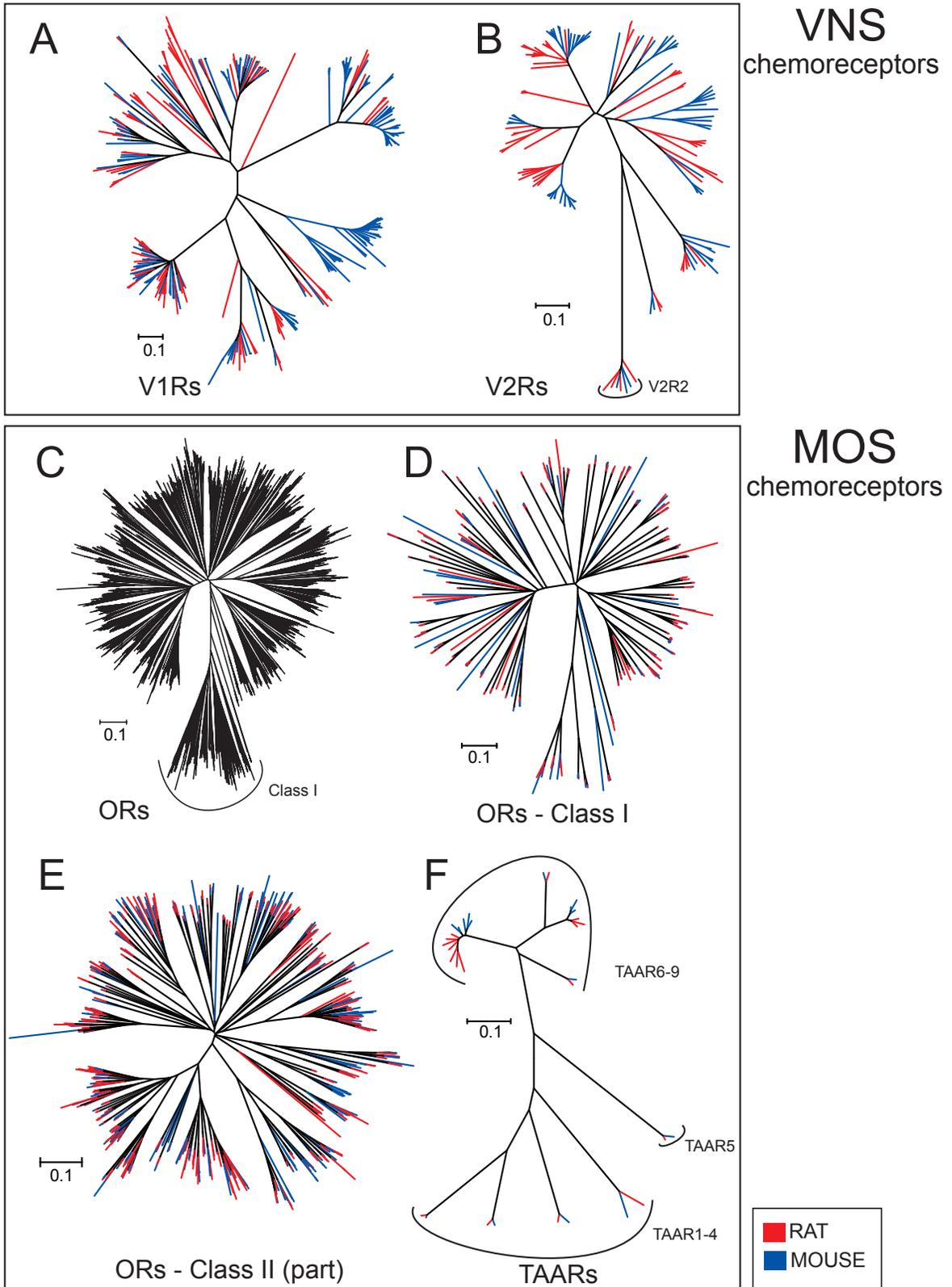


Figure 4.6 Unrooted phylogenetic trees of all putatively functional mouse and rat (A) V1Rs, (B) V2Rs, (C) ORs (D) Class I ORs, (E) Class II ORs from families 5 and 8 as defined by the HORDE database, and (F) TAARs. The trees were reconstructed using the neighbor-joining method with Poisson-corrected protein distances. The scale bars show 0.1 amino acid substitutions per site.

Figure 4.6



4.7 LITERATURE CITED

- Alioto, T. S., and J. Ngai. 2006. The repertoire of olfactory C family G protein-coupled receptors in zebrafish: candidate chemosensory receptors for amino acids. *BMC Genomics* **7**:309.
- Alioto, T. S., and J. Ngai. 2005. The odorant receptor repertoire of teleost fish. *BMC Genomics* **6**:173.
- Aloni, R., T. Olender, and D. Lancet. 2006. Ancient genomic architecture for mammalian olfactory receptor clusters. *Genome Biol* **7**:R88.
- Araneda, R. C., Z. Peterlin, X. Zhang, A. Chesler, and S. Firestein. 2004. A pharmacological profile of the aldehyde receptor repertoire in rat olfactory epithelium. *J Physiol* **555**:743-756.
- Baldwin, J. M. 1994. Structure and function of receptors coupled to G proteins. *Curr Opin Cell Biol* **6**:180-190.
- Baxi, K. N., K. M. Dorries, and H. L. Eisthen. 2006. Is the vomeronasal system really specialized for detecting pheromones? *Trends Neurosci* **29**:1-7.
- Belluscio, L., G. H. Gold, A. Nemes, and R. Axel. 1998. Mice deficient in G(olf) are anosmic. *Neuron* **20**:69-81.
- Benbernou, N., S. Tacher, S. Robin, M. Rakotomanga, F. Senger, and F. Galibert. 2007. Functional analysis of a subset of canine olfactory receptor genes. *J Hered* **98**:500-505.
- Chamero, P., T. F. Marton, D. W. Logan, K. Flanagan, J. R. Cruz, A. Saghatelian, B. F. Cravatt, and L. Stowers. 2007. Identification of protein pheromones that promote aggressive behaviour. *Nature* **450**:899-902.
- Dorries, K. M., E. Adkins-Regan, and B. P. Halpern. 1995. Olfactory sensitivity to the pheromone, androstenone, is sexually dimorphic in the pig. *Physiol Behav* **57**:255-259.
- Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**:783-791.
- Freitag, J., J. Krieger, J. Strotmann, and H. Breer. 1995. Two classes of olfactory receptors in *Xenopus laevis*. *Neuron* **15**:1383-1392.
- Gloriam, D. E., R. Fredriksson, and H. B. Schioth. 2007. The G protein-coupled receptor subset of the rat genome. *BMC Genomics* **8**:338.
- Glusman, G., A. Sosinsky, E. Ben-Asher, N. Avidan, D. Sonkin, A. Bahar, A. Rosenthal, S. Clifton, B. Roe, C. Ferraz, J. Demaille, and D. Lancet. 2000. Sequence, structure, and evolution of a complete human olfactory receptor gene cluster. *Genomics* **63**:227-245.
- Grus, W. E., P. Shi, and J. Zhang. 2007. Largest vertebrate vomeronasal type 1 receptor gene repertoire in the semiaquatic platypus. *Mol Biol Evol* **24**:2153-2157.
- Grus, W. E., P. Shi, Y. P. Zhang, and J. Zhang. 2005. Dramatic variation of the vomeronasal pheromone receptor gene repertoire among five orders of placental and marsupial mammals. *Proc Natl Acad Sci U S A* **102**:5767-5772.
- Grus, W. E., and J. Zhang. 2006. Origin and evolution of the vertebrate vomeronasal system viewed through system-specific genes. *BioEssays* **28**:709-718.
- Grus, W. E., and J. Zhang. 2004. Rapid turnover and species-specificity of vomeronasal pheromone receptor genes in mice and rats. *Gene* **340**:303-312.

- Halpern, M., Y. Daniels, and I. Zuri. 2005. The role of the vomeronasal system in food preferences of the gray short-tailed opossum, *Monodelphis domestica*. *Nutr Metab (Lond)* **2**:6.
- Hashiguchi, Y., and M. Nishida. 2007. Evolution of trace amine associated receptor (TAAR) gene family in vertebrates: lineage-specific expansions and degradations of a second class of vertebrate chemosensory receptors expressed in the olfactory epithelium. *Mol Biol Evol* **24**:2099-2107.
- Hoppe, R., T. D. Lambert, P. B. Samollow, H. Breer, and J. Strotmann. 2006. Evolution of the "OR37" subfamily of olfactory receptors: A cross-species comparison. *Journal of Molecular Evolution* **62**:460-472.
- Hudson, R., and H. Distel. 1983. Nipple Location by Newborn Rabbits - Behavioral Evidence for Pheromonal Guidance. *Behaviour* **85**:260-275.
- Katada, S., T. Hirokawa, Y. Oka, M. Suwa, and K. Touhara. 2005. Structural basis for a broad but selective ligand spectrum of a mouse olfactory receptor: mapping the odorant-binding site. *J Neurosci* **25**:1806-1815.
- Kelliher, K. R. 2007. The combined role of the main olfactory and vomeronasal systems in social communication in mammals. *Horm Behav* **52**:561-570.
- Kimoto, H., S. Haga, K. Sato, and K. Touhara. 2005. Sex-specific peptides from exocrine glands stimulate mouse vomeronasal sensory neurons. *Nature* **437**:898-901.
- Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**:150-163.
- Leinders-Zufall, T., P. Brennan, P. Widmayer, P. C. S, A. Maul-Pavicic, M. Jager, X. H. Li, H. Breer, F. Zufall, and T. Boehm. 2004. MHC class I peptides as chemosensory signals in the vomeronasal organ. *Science* **306**:1033-1037.
- Leinders-Zufall, T., A. P. Lane, A. C. Puche, W. Ma, M. V. Novotny, M. T. Shipley, and F. Zufall. 2000. Ultrasensitive pheromone detection by mammalian vomeronasal neurons. *Nature* **405**:792-796.
- Liberles, S. D., and L. B. Buck. 2006. A second class of chemosensory receptors in the olfactory epithelium. *Nature* **442**:645-650.
- Lindemann, L., M. Ebeling, N. A. Kratochwil, J. R. Bunzow, D. K. Grandy, and M. C. Hoener. 2005. Trace amine-associated receptors form structurally and functionally distinct subfamilies of novel G protein-coupled receptors. *Genomics* **85**:372-385.
- Malnic, B., J. Hirono, T. Sato, and L. B. Buck. 1999. Combinatorial receptor codes for odors. *Cell* **96**:713-723.
- Man, O., Y. Gilad, and D. Lancet. 2004. Prediction of the odorant binding site of olfactory receptor proteins by human-mouse comparisons. *Protein Sci* **13**:240-254.
- Mandiyani, V. S., J. K. Coats, and N. M. Shah. 2005. Deficits in sexual and aggressive behaviors in *Cnga2* mutant mice. *Nat Neurosci* **8**:1660-1662.
- Martini, S., L. Silvotti, A. Shirazi, N. J. Ryba, and R. Tirindelli. 2001. Co-expression of putative pheromone receptors in the sensory neurons of the vomeronasal organ. *J Neurosci* **21**:843-848.
- Mombaerts, P. 2004. Genes and ligands for odorant, vomeronasal and taste receptors. *Nat Rev Neurosci* **5**:263-278.

- Murphy, W. J., P. A. Pevzner, and S. J. O'Brien. 2004. Mammalian phylogenomics comes of age. *Trends Genet* **20**:631-639.
- Niimura, Y., and M. Nei. 2005. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proc Natl Acad Sci U S A* **102**:6039-6044.
- Niimura, Y., and M. Nei. 2007. Extensive gains and losses of olfactory receptor genes in Mammalian evolution. *PLoS ONE* **2**:e708.
- Pfister, P., J. Randall, J. I. Montoya-Burgos, and I. Rodriguez. 2007. Divergent Evolution among Teleost V1r Receptor Genes. *PLoS ONE* **2**:e379.
- Placyk, J. S., Jr., and B. M. Graves. 2002. Prey detection by vomeronasal chemoreception in a plethodontid salamander. *J Chem Ecol* **28**:1017-1036.
- Rat Genome Sequencing Consortium. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**:493-521.
- Restrepo, D., J. Arellano, A. M. Oliva, M. L. Schaefer, and W. Lin. 2004. Emerging views on the distinct but related roles of the main and accessory olfactory systems in responsiveness to chemosensory signals in mice. *Horm Behav* **46**:247-256.
- Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**:406-425.
- Sam, M., S. Vora, B. Malnic, W. Ma, M. V. Novotny, and L. B. Buck. 2001. Neuropharmacology. Odorants may arouse instinctive behaviours. *Nature* **412**:142.
- Sanz, G., C. Schlegel, J. C. Pernollet, and L. Briand. 2005. Comparison of odorant specificity of two human olfactory receptors from different phylogenetic classes and evidence for antagonism. *Chem Senses* **30**:69-80.
- Saraiva, L. R., and S. I. Korsching. 2007. A novel olfactory receptor gene family in teleost fish. *Genome Res* **17**:1448-1457.
- Scalia, F., and S. S. Winans. 1975. The differential projections of the olfactory bulb and accessory olfactory bulb in mammals. *J Comp Neurol* **161**:31-55.
- Schwenk, K. 1993. The evolution of chemoreception in squamate reptiles: a phylogenetic approach. *Brain Behav Evol* **41**:124-137.
- Shi, P., J. P. Bielawski, H. Yang, and Y. P. Zhang. 2005. Adaptive diversification of vomeronasal receptor 1 genes in rodents. *J Mol Evol* **60**:566-576.
- Shi, P., and J. Zhang. 2007. Comparative genomic analysis identifies an evolutionary shift of vomeronasal receptor gene repertoires in the vertebrate transition from water to land. *Genome Res* **17**:166-174.
- Silvotti, L., G. Giannini, and R. Tirindelli. 2005. The vomeronasal receptor V2R2 does not require escort molecules for expression in heterologous systems. *Chem Senses* **30**:1-8.
- Spehr, M., K. R. Kelliher, X. H. Li, T. Boehm, T. Leinders-Zufall, and F. Zufall. 2006a. Essential role of the main olfactory system in social recognition of major histocompatibility complex peptide ligands. *Journal of Neuroscience* **26**:1961-1970.
- Spehr, M., J. Spehr, K. Ukhanov, K. R. Kelliher, T. Leinders-Zufall, and F. Zufall. 2006b. Parallel processing of social signals by the mammalian main and accessory olfactory systems. *Cell Mol Life Sci* **63**:1476-1484.

- Stary, A., C. Suwattanasophon, P. Wolschann, and G. Buchbauer. 2007. Differences in (-)-citronellal binding to various odorant receptors. *Biochem Biophys Res Commun* **361**:941-945.
- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**:4876-4882.
- Wieland, K., H. M. Zuurmond, C. Krasel, A. P. Ijzerman, and M. J. Lohse. 1996. Involvement of Asn-293 in stereospecific agonist recognition and in activation of the beta 2-adrenergic receptor. *Proc Natl Acad Sci U S A* **93**:9276-9281.
- Xu, F., M. Schaefer, I. Kida, J. Schaefer, N. Liu, D. L. Rothman, F. Hyder, D. Restrepo, and G. M. Shepherd. 2005. Simultaneous activation of mouse main and accessory olfactory bulbs by odors or pheromones. *J Comp Neurol* **489**:491-500.
- Yang, H., P. Shi, Y. P. Zhang, and J. Zhang. 2005. Composition and evolution of the V2r vomeronasal receptor gene repertoire in mice and rats. *Genomics* **86**:306-315.
- Young, J. M., C. Friedman, E. M. Williams, J. A. Ross, L. Tonnes-Priddy, and B. J. Trask. 2002. Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Hum Mol Genet* **11**:535-546.
- Young, J. M., M. Kambere, B. J. Trask, and R. P. Lane. 2005. Divergent V1R repertoires in five species: Amplification in rodents, decimation in primates, and a surprisingly small repertoire in dogs. *Genome Res* **15**:231-240.
- Young, J. M., and B. J. Trask. 2007. V2R gene families degenerated in primates, dog and cow, but expanded in opossum. *Trends Genet*.
- Zhang, J. 2003. Evolution by gene duplication-an update. *Trends Ecol. Evol.* **18**:292-298
- Zhang, X., and S. Firestein. 2002. The olfactory receptor gene superfamily of the mouse. *Nat Neurosci* **5**:124-133.
- Zhang, X., I. Rodriguez, P. Mombaerts, and S. Firestein. 2004. Odorant and vomeronasal receptor genes in two mouse genome assemblies. *Genomics* **83**:802-811.

CHAPTER 5

ORIGIN OF THE VOMERONASAL SYSTEM IN THE COMMON ANCESTOR OF ALL VERTEBRATES

5.1 ABSTRACT

Comparative genomics provides a useful tool for inferring the evolutionary history and origin of physiological systems, particularly when this information is difficult to ascertain by morphological traits. One such example is the vomeronasal system (VNS), a vertebrate nasal chemosensory system that is responsible for detecting intraspecific pheromonal cues as well as environmental odorants. The morphological components of the VNS are found only in tetrapods, but the genetic components of the system have been found in teleost fish, in addition to tetrapods. To determine when the VNS originated, we searched for the VNS-specific genes in the genomes of two early diverging vertebrate lineages: the sea lamprey from jawless fishes and the elephant shark from cartilaginous fishes. Genes encoding V1Rs (vomeronasal type 1 receptors) and Trpc2 (transient receptor potential cation channel, subfamily C, member 2), two components of the vomeronasal signaling pathway, are present in the sea lamprey genome and both are expressed in the olfactory epithelium, suggesting the presence of a primordial VNS in the common ancestor of all extant vertebrates. Additionally, all three VNS genes, Trpc2, V1Rs, and V2Rs (vomeronasal type 2 receptors), are found in the genome of the elephant shark. Coupled with evolutionary analysis of the vertebrate main

olfactory system and taste system, our study reveals staggered origins within and between vertebrate sensory systems. These results are important for understanding how vertebrate sensory systems originated and illustrate the utility of genome sequences of early diverging vertebrates for understanding the evolution of vertebrate-specific systems

5.2 INTRODUCTION

Comparative genomics is a useful tool for understanding the evolution of physiological systems (Arendt 2003; Zhang and Webb 2003; Okabe and Graham 2004; Go et al. 2005; Serb and Oakley 2005; Grus and Zhang 2006), because it provides information about the phylogenetic distribution of system-specific genetic components (Wray and Abouheif 1998; Serb and Oakley 2005). This approach allows a more complete understanding of the evolution of physiological systems than was previously available with only morphological characterization. For example, this approach has been used to identify a common origin for the invertebrate and vertebrate visual systems, two systems that were thought to have distinct origins (Gehring 1996; Arendt et al. 2004), and to identify how lepidopteran scales and the tetrapod parathyroid gland evolved in their respective lineages (Galant et al. 1998; Okabe and Graham 2004). Here, we use this approach to determine the origin of the vertebrate vomeronasal system (VNS), a nasal chemosensory system responsible for detecting intraspecific pheromonal cues as well as some environmental odorants.

The VNS is one of the two nasal chemosensory systems found in many vertebrates (the other being the main olfactory system) (Grus and Zhang 2006). Prior to the identification of the VNS genetic components, the evolutionary history of the VNS

was based on the phylogenetic distribution of VNS morphological characteristics, the vomeronasal organ (VNO), an organ in the nasal cavity with sensory neurons expressing vomeronasal receptors, and the accessory olfactory bulb, which is the part of the brain excited by the VNO sensory neurons (Bertmar 1981; Eisthen 1992; Dulka 1993; Eisthen 1997). Because these morphological characters are found only in tetrapods, the VNS was thought to have evolved in the common ancestor of tetrapods as an adaptation to terrestrial living (Bertmar 1981). However, finer-scale studies of the morphological components, such as sensory neuron morphology, gave hints that the VNS might exist in an unrecognized form in teleost fish (Eisthen 1992; Dulka 1993). Furthermore, evidence of VNS development in larval amphibians and neotenic salamanders suggests that the VNS is not an adaptation to terrestrial life (Eisthen 2000; Jermakowicz et al. 2004). Therefore, an additional type of character became necessary to address the lingering possibility of the presence of the VNS in teleosts.

The solution came with the identification of VNS-specific genes, which are defined as the genetic components of the vomeronasal signal transduction pathway that are not used for other functions and include two receptor family genes and a channel protein gene. The two families of VNS-specific receptors, V1Rs and V2Rs, were first identified from rodents (Dulac and Axel 1995; Herrada and Dulac 1997; Matsunami and Buck 1997; Ryba and Tirindelli 1997). Both families are 7 transmembrane domain G protein-coupled receptors (GPCRs), but they are evolutionarily unrelated. Each type of receptor couples to a unique G protein (Dulac and Axel 1995; Ryba and Tirindelli 1997), and are spatially segregated in expression in tetrapods (Hagino-Yamagishi et al. 2004; Takigami et al. 2004; Date-Ito et al. 2008). It has been proposed that V1Rs and V2Rs are

specialized for binding to air-borne molecules and water-soluble peptides, respectively, although this functional separation may not be complete (Shi and Zhang 2007). In addition to the receptors, a calcium channel protein, Trpc2, was identified with specific expression in the VNO (Liman, Corey, and Dulac 1999). Its indispensable and sole function in the vomeronasal signal transduction pathway is supported by studies of Trpc2-deficient mice, which show altered intraspecific interactions such as sex discrimination and male-male aggression (Leypold et al. 2002; Stowers et al. 2002; Kimchi, Xu, and Dulac 2007) and by evolutionary studies that found catarrhine primates, which do not have a VNO, have a nonfunctional Trpc2 gene (Liman and Innan 2003; Zhang and Webb 2003).

Interestingly, despite that the morphological traits of the VNS are found only in tetrapods, VNS-specific genes are present in teleost fish and exhibit expression patterns consistent with their involvement in a distinct teleost olfactory system (Cao, Oh, and Stryer 1998; Naito et al. 1998; Dukes et al. 2004; Hansen, Anderson, and Finger 2004; Pfister and Rodriguez 2005; Sato, Miyasaka, and Yoshihara 2005). These findings suggest that a primordial VNS existed in the common ancestor of teleosts and tetrapods (Grus and Zhang 2006). Because earlier diverging vertebrates, such as jawless fish and cartilaginous fish, have a single olfactory organ as teleosts do, the VNS was not thought to be present in these lineages. However, based on the findings that teleost fish have one organ but two distinct olfactory signaling pathways, it is possible that even earlier diverging vertebrates may possess the VNS. Therefore, armed with the three types of VNS-specific genes and the new whole-genome sequences of a jawless fish (sea lamprey)

and a cartilaginous fish (elephant shark), we explore the evolutionary history of the VNS in the common ancestor of all vertebrates.

5.3 MATERIALS AND METHODS

5.3.1 Computational identification of sea lamprey, elephant shark, and tunicate olfaction genes

TBLASTN searches for V1R, V2R, and *Trpc2* genes were conducted on the sea lamprey (*Petromyzon marinus*), elephant shark (*Callorhinchus milii*), tunicate (*Ciona intestinalis*), and amphioxus (*Branchiostoma floridae*) genomes. The 5.9× coverage sea lamprey genome sequence is available from Ensembl (http://pre.ensembl.org/Petromyzon_marinus/index.html). The 1.4× elephant shark genome sequence is available from the Institute of Molecular and Cellular Biology (<http://esharkgenome.imcb.a-star.edu.sg/>) (Venkatesh et al. 2007). The 11× tunicate genome sequence is available from the Joint Genome Institute (<http://genome.jgi-psf.org/Cioin2/Cioin2.home.html>). The 8.1× amphioxus genome sequence is available from the Joint Genome Institute (<http://genome.jgi-psf.org/Braf11/Braf11.home.html>). Zebrafish V1Rs, V2Rs, and *Trpc2* (Pfister and Rodriguez 2005; Sato, Miyasaka, and Yoshihara 2005; Shi and Zhang 2007) were used as query sequences. Because these taxa are distantly related to zebrafish, we used an e-value cut-off of 10^{-1} . The hits were then used as queries to BLAST the nr database of GenBank. A putative V1R gene was considered to be real if its best hit was a previously known V1R, and similar criteria were used for *Trpc2* and V2Rs.

5.3.2 Sequence alignment and phylogenetic analysis

Gene sequences were aligned per protein sequence alignment by Clustal_X (Thompson et al. 1997) with manual adjustment. Phylogenetic trees were reconstructed using the neighbor-joining method (Saitou and Nei 1987) with Poisson-corrected distances (Nei and Kumar 2000), and were evaluated by 1000 bootstrap replications (Felsenstein 1985).

5.3.3 Identification of expression pattern of sea lamprey olfaction genes

Adult sea lampreys were obtained from Hammond Bay Biological Station (Millersberg, Michigan). They were euthanized according to university animal care procedures (UCUCA #09470). Genomic DNA was isolated from fresh tissue with a DNeasy Tissue Kit (Qiagen). Sea lamprey tissues (testes, heart, tongue, olfactory organ) were dissected out and frozen in an ethanol/dry ice bath. RNAqueous-4PCR (Ambion) was used to extract RNA from the tissues and RetroScript kit (Ambion) was used for cDNA synthesis.

Putative sea lamprey Trpc2 and V1R genes were amplified from both genomic DNA and cDNA. Based on an alignment of Trpc2 from zebrafish, frog, mouse, and dog, nested degenerate primers for Trpc2 were designed using CODEHOP (Rose et al. 1998; Rose, Henikoff, and Henikoff 2003). These primers amplify a portion corresponding to mammalian Trpc2 exons 13-15. With primers 1315F1 and 1315R1 (Table 5), the following cycling conditions were used: 94°C for 3 minutes, followed by 35 cycles of 94°C for 30 seconds, 50°C for 30 seconds, and 72°C for 1 minute, followed by a final elongation for 5 minutes at 72°C. The nested PCR with primers 1315F2 and 1315R2

(Table 5) was conducted with the same cycling conditions except that the annealing temperature was increased to 53°C. V1Rs have a single coding exon, making it difficult to distinguish true V1R expression from genomic contamination of the cDNA. To investigate V1R expression patterns unambiguously, we performed 3'RACE with a FirstChoice RLM RACE kit (Ambion) with primers V1RRACEOUT and V1RRACEIN, following the manufacturers instructions. All PCRs were carried out in a 25 uL volume with the following final concentrations: 40nM of each primer, 200uM dNTPs, and 1.5mM MgCl₂. PCR amplicons were confirmed on an agarose gel.

5.4 RESULTS

5.4.1 V1Rs from early diverging vertebrates

We were able to identify three and two putatively functional V1R genes from the sea lamprey and elephant shark genomes, respectively. Given the relatively high coverage of the sea lamprey genome (5.9×), this likely represents the majority of the lamprey V1R repertoire. However, given the extreme protein distance between the known teleost V1Rs and the putative lamprey V1Rs and the rapid evolution characteristic of this gene family in mammals (Grus and Zhang 2006), we cannot exclude the possibility that some lamprey V1Rs were too divergent to detect by our methods. The low coverage of the elephant shark genome (1.4×) suggests that the V1R repertoire might be incomplete based on this search. The small sea lamprey and elephant shark V1R repertoire size is similar to what is observed in teleost fish (Saraiva and Korsching 2007), but contrasts to that in most lineages of tetrapods (Shi and Zhang 2007).

To understand the evolution of the V1Rs in early diverging vertebrates, we reconstructed a protein-sequence-based V1R phylogeny with the 3 putatively functional sea lamprey V1Rs, 2 putatively functional elephant shark V1Rs, 6 putatively functional zebrafish V1Rs (Saraiva and Korsching 2007), 5 putatively functional Tetraodon (*Tetraodon nigriviridis*) V1Rs (Saraiva and Korsching 2007), 21 putatively functional frog (*Xenopus tropicalis*) V1Rs (Shi and Zhang 2007), and 13 representative mammalian V1Rs (Shi et al. 2005) (Fig. 5.1A). These 13 mammalian genes represent 13 V1R gene groups present in the mouse and rat genomes (Grus et al. 2005). Although other placental mammal V1R families likely exist and marsupial and monotreme V1R families are unique to their lineages, previous studies of V1Rs from non-mammalian vertebrates have shown that all mammalian V1Rs will cluster together in relation to non-mammalian V1Rs (Grus et al. 2007; Saraiva and Korsching 2007). One sea lamprey V1R and one elephant shark V1R cluster with a teleost V1R clade that was lost in mammals: teleost V1R3-4. These teleost V1Rs are unique in that their coding regions contain introns (Saraiva and Korsching 2007). However, according to the BLAST searches, the sea lamprey and elephant shark V1R3-4s contain a single coding exon, suggesting that these teleost V1Rs acquired the introns after the teleost-tetrapod separation. The other elephant shark V1R clusters with teleost V1R2, which has orthologs in all teleost fish thus far studied (Pfister et al. 2007; Saraiva and Korsching 2007). The remaining sea lamprey V1Rs do not cluster with other teleost V1Rs. Neither elephant shark nor sea lamprey V1Rs cluster with the teleost-specific V1R5-6 genes or the V1R1 gene, which has homologs in tetrapods. If we assume the V1R tree is rooted with T2Rs (bitter taste receptors), the phylogenetic reconstruction (Fig. 5.1A) strongly supports the clustering of

both sea lamprey and elephant shark genes with other. The tree also presents strong evidence that the ancestral vertebrate genome had at least one V1R, while the ancestral jawed vertebrate genome likely had at least two V1Rs. With 3'RACE, we confirmed that at least one of the sea lamprey V1Rs is expressed in the olfactory epithelium (Fig. 5.1B). Although there are 3'RACE V1R products from cDNA from testes and tongue, these tissues do not express *Trpc2* (see below), so their expression does suggest the VNS pathway is acting in non-chemosensory tissue in early diverging vertebrates.

5.4.2 V2Rs from early diverging vertebrates

No V2R homologs were found in the sea lamprey genome. However, V2Rs were present in the elephant shark genome. Because of their complex multi-exon structure and because the elephant shark genome was not assembled, no complete V2Rs were identified. However, both the 7 transmembrane domain and extracellular domain of the V2R have multiple hits, and two shark V2Rs contain multiple exons and cover over 600 amino acids each. To recover as many as possible unique shark V2Rs, we required that a hit be at least 200 amino acids long in the 7 transmembrane region and share no more than 1% sequence identity with another elephant shark V2R in this region. Based on these criteria, we found 32 unique shark V2Rs, which is likely a conservative estimate for the total number of shark V2Rs. To understand the evolutionary relationships of the shark V2Rs with other known vertebrate V2Rs, we analyzed the 25 V2Rs of these 32 partial genes with the most complete sequence in the 7 transmembrane domain with 43 zebrafish V2Rs, 18 fugu V2Rs, and 4 tetraodon V2Rs, 249 frog V2Rs, and 5 representative mammalian V2Rs (Shi and Zhang 2007). These mammalian V2R genes

were chosen from diverse mammalian V2R clades including the V2R2 clade. As in the V1Rs, teleost V2Rs form distinct clades from tetrapod V2Rs, aside from the V2R2 group (Shi and Zhang 2007). The 25 shark V2Rs used range in length from 222 to 684 amino acids. Among these shark V2Rs was an ortholog to the V2R2 clade that was previously known to have distinct expression, function, and origin from other V2Rs (Shi and Zhang 2007; Young and Trask 2007). All of the remaining elephant shark V2Rs fall within the teleost fish V2R clade. The majority of these shark V2Rs (16 of 25) cluster together with high bootstrap support in a shark-specific clade. One of the remaining shark V2Rs appears to be orthologous to a Fugu V2R. The remaining seven shark V2Rs are found within the teleost V2R clade, but cluster with high support with at most one other elephant shark V2R.

5.4.3 Trpc2 from early diverging vertebrates

From the sea lamprey genome, we identified two copies of Trpc2 exons 14 and 15. One of these contigs has a nonsense mutation in exon 14 and appears to be a pseudogene. The other contig has an open reading frame across this region. Additionally, we amplified a 554 nucleotide cDNA sequence corresponding to Trpc2 exons 12-15. The 3' end of this sequence matched the genomic sequence containing exons 14 and 15 with the open reading frame. In addition to the two copies of exons 14 and 15, we identified a single copy of exon 7 from the genome sequence, corresponding to a 92 amino acid region unique to the longest exon in vertebrate Trpc2. However, BLAST searches of the sea lamprey genome did not identify contigs with high sequence

similarity to exons 12 or 13, or any of the other exons. From the elephant shark genome we identified one sequence that corresponded to vertebrate Trpc2 exon 8.

To confirm that the sequences we identified are indeed the Trpc2 gene, we reconstructed a phylogeny including the sea lamprey Trpc2 exons 12-15 with the homologous regions from tetrapod and teleost Trpc2 protein sequences and other mouse members of the Trpc gene family (Fig. 5.3A). The identified sea lamprey Trpc2 sequence clusters with high bootstrap with the other vertebrate Trpc2 genes, strongly suggesting that this sequence comes from the sea lamprey ortholog of Trpc2. We did not include the sea lamprey Trpc2 exon 7 sequence in the alignment for the phylogeny because we do not know which copy it belongs with. Because we could not identify the same exons of Trpc2 from both the lamprey and shark genomes, we could not include both species in the same tree. However, a separate phylogenetic reconstruction using exon 8 sequences confirms that the partial Trpc2 sequence of the elephant shark is indeed Trpc2 (Fig. 5.3B). Note that the gene tree including the elephant shark Trpc2 exon 8 is not completely concordant with the known species tree, as the shark appears sister to the frog in the gene tree. This discordance may have resulted from the use of a short sequence (67 amino acids) in a region that has high sequence identity between teleosts and tetrapods. Regardless of cause of the discordance, the shark sequence clusters with the other vertebrate Trpc2 sequences, rather than with other Trpc sequences. To determine if the sea lamprey Trpc2 functions in chemoreception, we examined the expression pattern of Trpc2 in sea lamprey using RT-PCR. We found that the sea lamprey Trpc2 is expressed in the olfactory tissue and has low expression in the heart, but it is not expressed in the testes or tongue (Fig. 5.3C).

5.4.4 VNS-specific genes are absent from non-vertebrate chordate genomes

To investigate if these genes were also present in non-vertebrate chordates, we searched for them in two representative genomes from closely related non-vertebrate chordate lineages, urochordates and cephalochordates. None of the three VNS-specific genes were found in either the tunicate or the amphioxus genomes. The absence of the three VNS-specific genes from the genomes of representative species from two distinct non-vertebrate chordate lineages, although not definitive, strongly suggests that this signal transduction pathway originated in the most recent common ancestor of all extant vertebrates.

5.5 DISCUSSION

Previous studies have established that the VNS likely existed in the common ancestor of bony vertebrates. To examine if the system originated even earlier, we examined the elephant shark and sea lamprey, which represent cartilaginous fish and jawless vertebrates, respectively. With both computational and experimental methods, we identified two of the three VNS-specific genes from the sea lamprey (V1Rs and Trpc2) and all three VNS-specific genes (V1Rs, V2Rs, and Trpc2) from the elephant shark. In addition, we showed that lamprey V1Rs and Trpc2 are expressed exclusively in the olfactory epithelium, suggesting an ancestral role for these genes in chemoreception. Although V2Rs were identified from the elephant shark genome, they were not found in the sea lamprey genome. Three scenarios could cause this result. First, V2Rs might not be present in the sea lamprey genome, either because of gene loss in the lamprey lineage

or because this gene superfamily did not originate until after the divergence of jawed and jawless vertebrates. Second, V2Rs might be present in the sea lamprey genome but absent in the genome assembly because the current assembly is incomplete. Given that we could only identify three exons of *Trpc2* from the genome sequence, this explanation seems plausible. However, even if V2Rs are present in the sea lamprey genome, the gene family would be very small because the elephant shark genome sequence with much lower coverage has over 30 V2Rs. Finally, V2Rs in the sea lamprey genome could be so divergent from the other vertebrate V2Rs that they are not hit in our BLAST search. Attempts to amplify V2Rs from the sea lamprey genomic DNA or olfactory cDNA with degenerate V2R primers (Cao, Oh, and Stryer 1998) were also unsuccessful. However, the divergent sequence hypothesis cannot explain our BLAST results, as we were able to identify related non-V2R genes, such as calcium-sensing receptors (CaSRs) and metabotropic glutamate receptors. Thus, our search was sensitive enough to identify distantly related sequences and should have been able to identify divergent V2Rs. While we are unsure of the exact time of origin of the V2R superfamily, it arose at least before the divergence of cartilaginous fish and bony vertebrates. Identifying these genes from the other jawless vertebrate lineage, the hagfishes, would reveal if this family was present in the ancestor of all extant vertebrates and was subsequently lost in the lamprey lineage.

Although the evolutionary history of V2Rs remains ambiguous, the two other VNS-specific genes, V1Rs and *Trpc2*, are clearly present in the sea lamprey genome and must have been present in the common ancestor of all extant vertebrates. Additionally, the only sea lamprey tissue with expression of both V1Rs and *Trpc2* is the olfactory epithelium suggesting an ancestral role for this system in chemoreception. Further

support for the origin of the VNS in the ancestor of vertebrates would come from identifying the exact expression location within the sea lamprey olfactory epithelium. Like in teleost fish, olfactory sensory neurons of sea lamprey are polymorphic, with three morphologies: ciliated, microvillar, and crypt (Laframboise et al. 2007). In teleost fish, the VNS-specific genes have been shown to be expressed in the apical layer of microvillar sensory neurons while the main olfactory system (MOS)-specific genes are expressed in a more basal layer of ciliated sensory neurons (Hansen, Anderson, and Finger 2004; Sato, Miyasaka, and Yoshihara 2005). If the two olfactory systems are distinct in the sea lamprey, we would predict that such spatial distinction is seen in the expression of sea lamprey V1Rs and Trpc2 compared to main olfactory system genes, which will be interesting to confirm in the future.

While the expression patterns of the VNS-specific genes suggest a role in chemoreception, it is still unclear what the exact physiological function of the VNS was in early diverging vertebrates. In tetrapods, the VNS was thought to be the olfactory system for detecting pheromones, while the main olfactory system detected general odorants (Scalia and Winans 1975). However, experimental evidence suggests that there is not such a clear functional distinction (Restrepo et al. 2004; Baxi, Dorries, and Eisthen 2006; Spehr et al. 2006; Kelliher 2007). Sea lampreys produce unique bile acids which act as pheromones both in migration and mate finding (Li, Sorensen, and Gallaher 1995; Li et al. 2002; Siefkes and Li 2004). However, bile acids in teleost fish are known to require components of the main olfactory signal transduction pathway (Hansen et al. 2003), and interruption of the VNS signal transduction pathway had no effect on bile acid olfactory response (Hansen et al. 2003). In contrast to the large V1R gene repertoire and

rapid gene turnover in mammals, the small repertoire size and strict orthologous relationships in this gene family across teleosts suggest that VIR chemoreception may play different physiological roles in teleosts than in mammals. The evolutionary patterns observed in the early diverging vertebrate VIRs are similar to what is seen in teleosts, suggesting that the sea lamprey and elephant shark VIRs might have a function similar to those in teleosts.

Given that the VNS-specific genetic components arose in the ancestor of vertebrates, how did the system originate? Because of their related functions, it was hypothesized that the VNS arose via duplication of the main olfactory system (Eisthen 1992; Eisthen 1997). While the two systems are morphologically similar and, in tetrapods, the two distinct organs both develop from the olfactory placode (Taniguchi et al. 1996; Taniguchi and Taniguchi 2008), the genetic components of their signaling pathways are distinct and non-homologous (Dulac and Axel 1995). The two chemoreceptor families expressed in teleost and tetrapod main olfactory systems, odorant receptors (ORs) and trace amine-associated receptors (TAARs), are not closely related to the VIRs and V2Rs. In contrast, both vomeronasal receptor families have closely related homologs in the vertebrate taste system: VIRs are closely related to T2R bitter taste receptors (Adler et al. 2000) and V2Rs are closely related to T1R sweet and umami taste receptors (Hoon et al. 1999). Thus, it is possible that the VNS arose as a duplication of the taste system (or the converse). However, BLAST searches of the sea lamprey genome with either T1Rs or T2Rs as query sequences reveal no taste receptors, suggesting that these genes are not present in the sea lamprey genome. Additionally, the sea lamprey genome does not appear to have an ortholog for Trpm5, a channel protein

necessary for bitter, sweet, and umami taste signal transduction (Perez et al. 2002). The taste receptors and Trpm5 are also absent from the tunicate genome. Because jawless fish do not have a tongue that is homologous to the tetrapod tongue (Iwasaki 2002), it might be expected that these genes are absent. However, this organ is also absent from jawed fish (Iwasaki 2002), while these animals have both T1Rs and T2Rs (Shi and Zhang 2006). Additionally, searches of the elephant shark genome reveal orthologs of T1Rs and Trpm5 (Fig. 5.4), but no T2R homologs. Since the VNS arose prior to the taste system, it does not appear that the VNS evolved as the result of a duplication of the taste system, although the converse remains a possibility.

In contrast to the vertebrate-specific VNS and taste system, the vertebrate main olfactory system appears to be much older. The two MOS-specific receptor families, ORs and TAARs, have been identified with olfactory epithelium expression in lamprey (Berghard and Dryer 1998; Freitag et al. 1999). Additionally, a closely related receptor has been identified in the amphioxus genome and shows expression in amphioxus primary sensory neurons (Satoh 2005). Further BLAST searches on the amphioxus genome sequence identify that this gene is part of a large receptor family that is homologous to vertebrate ORs (Grus and Zhang, *unpublished data*), suggesting that ORs originated in invertebrates. Interestingly, this gene family is absent from the tunicate genome (Satoh 2005). This combined analysis of vertebrate chemosensory receptors has unveiled a staggered pattern for the timing of vertebrate sensory system origins (Fig. 5.5). The two main olfactory system chemoreceptor families, ORs and TAARs, arose in the common ancestor of cephalochordates and vertebrates and in the ancestor of vertebrates, respectively. The two VNS chemoreceptor families, V1Rs and V2Rs, evolved in the

ancestor of vertebrates and jawed vertebrates, respectively. The two taste chemoreceptor families, T1Rs and T2Rs, evolved in the ancestor of jawed vertebrates and bony vertebrates, respectively. Searching for T2Rs in non-teleost ray-finned fish, such as the sturgeon or the bichir, would give a clearer picture of T2R origin. This staggered pattern could reflect changes in environment, social structure, ecological exploitations or alternative sensory systems, such as electroreception or the lateral line system, that appear in different lineages throughout vertebrate evolution.

Recent comparative analysis of vertebrate and non-vertebrate genomes suggests that 22% of vertebrate genes have no homologs in non-vertebrates (Prachumwat and Li 2008). However, since this analysis includes only representative vertebrate genomes from tetrapods and teleosts, it does not tell us when in vertebrate evolution these vertebrate-specific genes arose. Although the current sea lamprey genome assembly has limited coverage and the shark genome sequence has low coverage (Venkatesh et al. 2007), we have shown that these genome sequences provide a more complete picture of vertebrate genome evolution. The majority of identified vertebrate-specific genes were not part of large gene families, but an exception was observed for genes with a biological process of response to external biotic and abiotic stimuli and with cellular localization to extracellular space and plasma membrane (Prachumwat and Li 2008). These characteristics exactly describe chemosensory receptors and immunity genes. Along these lines, V2Rs and T1Rs are not the first identified gene families that have evolved early in vertebrate evolution but after the divergence of the jawless vertebrates. In addition, the adaptive immune system found in tetrapods and teleosts evolved in the common ancestor of jawed vertebrates (Bartl et al. 1997), while jawless vertebrates have

independently evolved a different adaptive immune system (Alder et al. 2005; Pancer et al. 2005). Thus, characteristics of vertebrate-specific genes suggest specific physiological functions whose evolutionary origins can be determined by focusing on early diverging vertebrate genomes such as those of jawless vertebrates and cartilaginous fish.

5.6 ACKNOWLEDGMENTS

We thank Drs. Soochin Cho, Jerry Smith, and Sean Parghi for technical assistance. We thank Tom Finger, Anne Hansen, Nicole Shultz, and Jennell Barrows at the University of Colorado Health Sciences Center for histological expertise and advice. We thank the Hammond Bay Biological Station in Millersburg, MI, for sea lampreys. This work was supported by National Institutes of Health (NIH) grant GM80285 to J.Z. W.E.G was supported by NIH training grant T32HG000040, Rackham One-Term Dissertation Grant, and Rackham Graduate School Research Grant.

Figure 5.1 V1Rs are found in both the sea lamprey and elephant shark genomes and are expressed in the sea lamprey olfactory epithelium. (A) Unrooted phylogenetic reconstruction of 3 sea lamprey, 2 elephant shark, 7 teleost fish from zebrafish and fugu, 21 frog V1Rs, and 13 mammalian V1Rs. If we assume T2R taste receptors (black) are the outgroup, all vertebrate V1Rs cluster together. Bootstrap support for each of the four teleost V1R families is given. Scale bar shows 0.2 amino acid substitutions per site. (B) Sea lamprey V1R expression in the olfactory epithelium. Presence of cDNA in all samples was verified by amplification of sea lamprey actin.

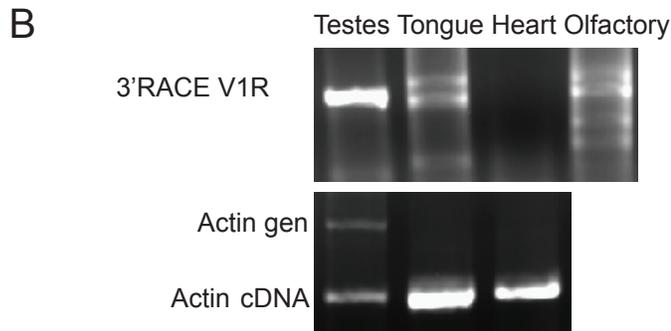
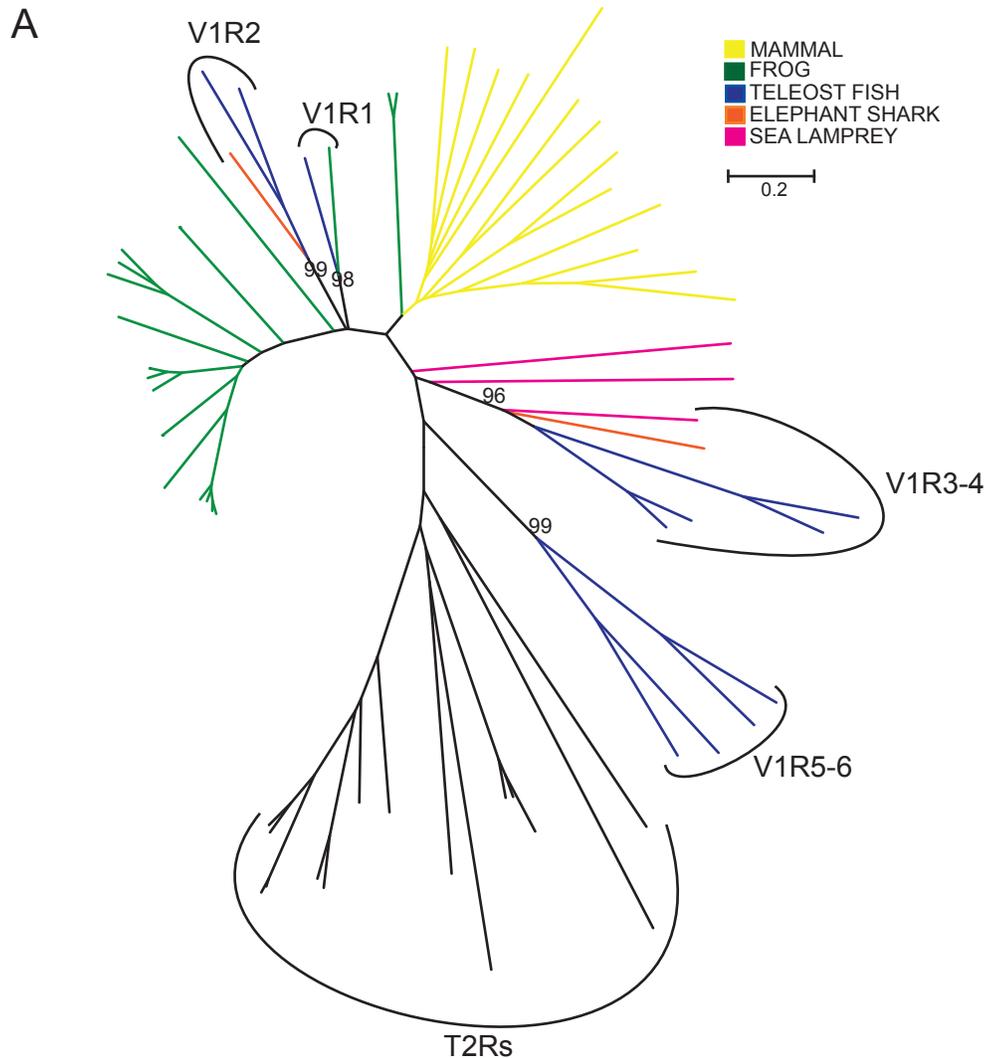


Figure 5.2 V2Rs are present in the elephant shark genome. An unrooted phylogenetic reconstruction of the 7 transmembrane domain from 5 mammalian V2Rs, 249 frog V2Rs, 43 zebrafish V2Rs, 4 tetraodon V2Rs, 18 fugu V2rs, and 25 elephant shark V2Rs. The scale bar shows 0.1 amino acid substitutions per site. The unique subfamily V2R2 is shown and includes an elephant shark V2R.

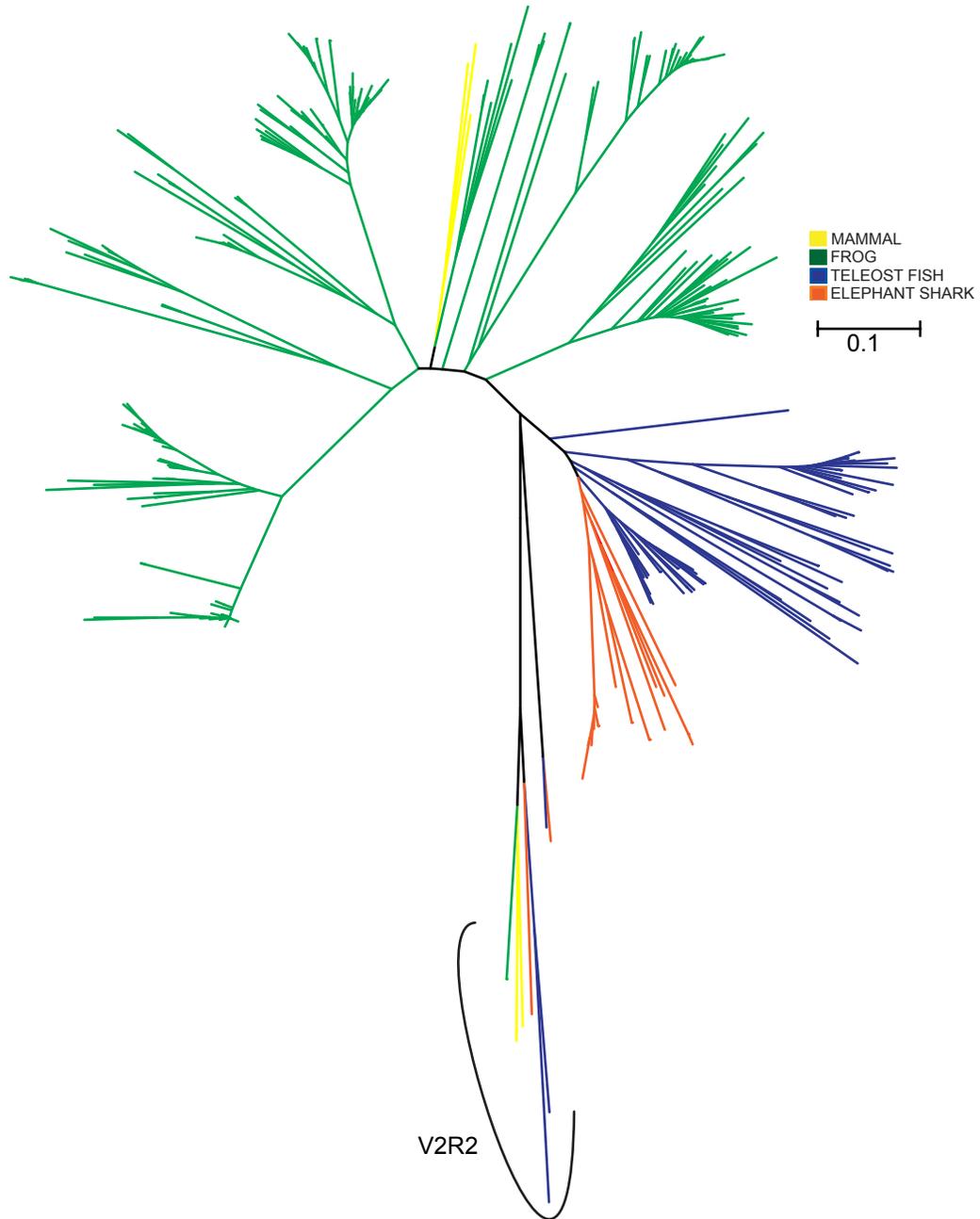


Figure 5.3 Trpc2 is present in the sea lamprey and elephant shark genomes. Phylogenetic reconstruction of (A) sea lamprey Trpc2 exons 7, 14, and 15 and (B) elephant shark Trpc2 exon 8 with Trpc2 from mouse, rat, tamarin, fugu, zebrafish, and frog. The trees are rooted with Trpc1, Trpc3, Trpc4, Trpc5, Trpc6, and Trpc7 from mouse. The scale bar shows 0.1 amino acid substitutions per site. (C) Trpc2 is expressed in the sea lamprey olfactory epithelium. Presence of cDNA in all samples was verified by amplification of sea lamprey actin (Figure 5.1B).

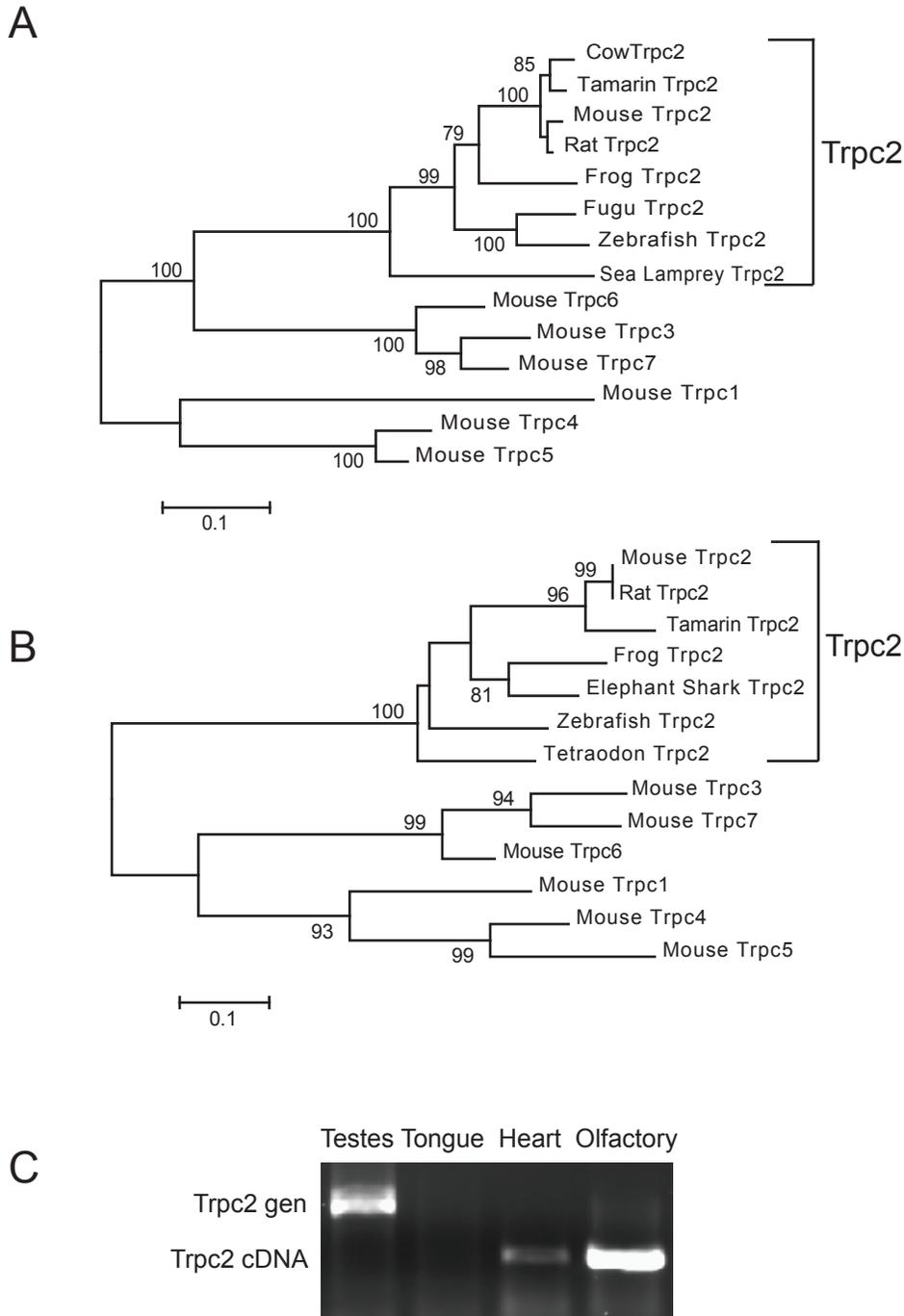


Figure 5.4 Genetic components of the taste system are present in the elephant shark genome. **(A)** Trpm5 and **(B and C)** T1Rs are found in the genome. A 342-amino-acid sequence from elephant shark Trpm5 is aligned with Trpm5 from mouse, rat, chimpanzee, dog, zebrafish, and tetraodon. The tree is rooted with other members of the Trpm family from mouse. Although no full length T1R was identified from the elephant shark genome, two elephant shark sequences hit the **(B)** N-terminal and **(C)** C-terminal ends of the T1R amino acid sequences. The elephant shark T1Rs were aligned with T1Rs from (Shi and Zhang 2006). From the trees it is difficult to classify either the N-terminal or C-terminal elephant shark T1Rs into the predefined T1R1, T1R2, or T1R3 genes, and it is unclear whether the four elephant shark partial T1Rs represent two, three, or four genes. The N-terminal tree (B) suggests at least one elephant shark T1R is homologous to T1R3, a member of the gene family that will heterodimerize with either T1R1 or T1R2 for umami and sweet taste reception in mammals, respectively. The C-terminal tree (C) suggests the duplication in the T1R family leading to T1R1 and T1R2 happened in the ancestor of bony vertebrates. Both trees offer different relationships between the three types of T1Rs, but the N-terminal tree (C) is more consistent with previous studies of the evolution of T1Rs in teleosts and tetrapods (Shi and Zhang 2006). Additionally, neither T1R tree contains a monophyletic T1R2 clade, as the mammalian T1R2s are distinct from the teleost fish T1R2s. This finding is not grossly inconsistent with the previous evolutionary study, which showed only marginal support for a clade containing both mammalian and tetrapod T1R2s (Shi and Zhang 2006). Given that the teleost fish T1R2/T1R3 complex responds to amino acids (Oike et al. 2007) rather than sweet tastants as in mammals (Mombaerts 2004), we do not know if the polyphyly and paraphyly of T1R2s in B and C, respectively, suggests divergent sequences or truly independent origins for teleost and mammalian T1R2s. More complete sequence of these four partial elephant shark T1Rs would reveal more about the evolution of this family and the T1R1/T1R3 and T1R2/T1R3 complexes. The scale bars show 0.1 amino acid substitutions per site. The T1R trees are rooted with vertebrate V2Rs. Bootstrap support over 70% is shown.

Figure 5.4

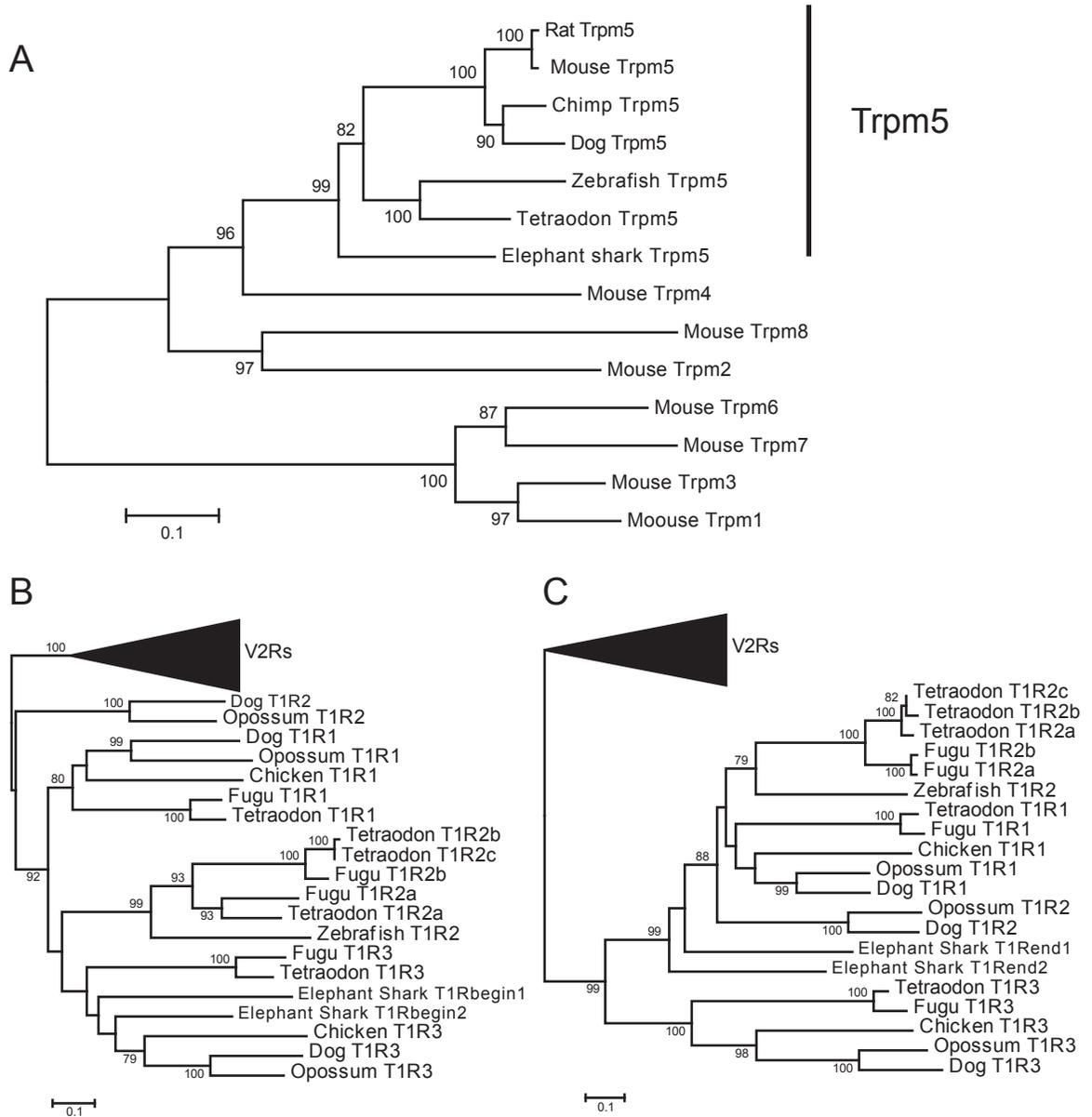


Figure 5.5 Origins of vertebrate chemosensory systems are staggered throughout vertebrate evolution.

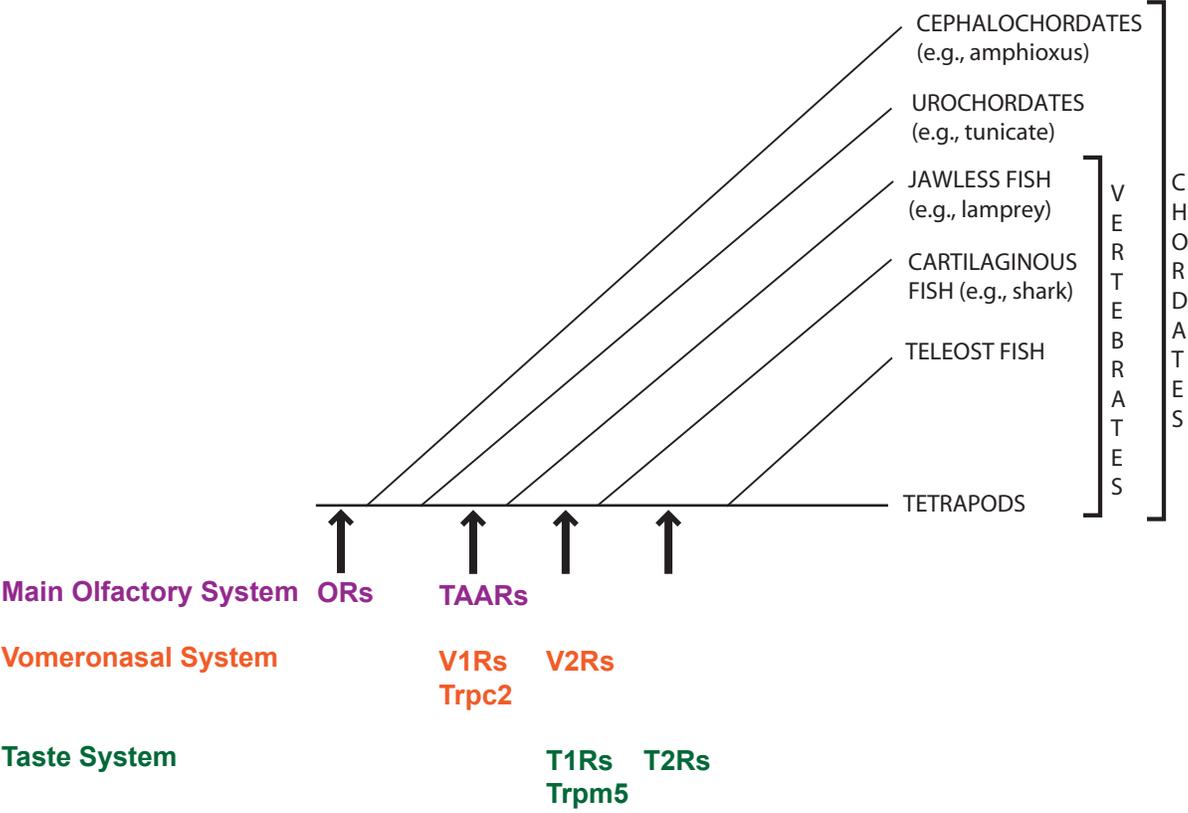


Table 5.1 Primers for sea lamprey VNS genes

Gene	Primers	Sequence (5' -> 3')	Annealing Temperature (°C)
V1R	V1RRACEOUT	CTGAGCGTCTTTCGCTTCTT	55
	V1RRACEIN	GTCACCAACAAGGCGTACCT	55
Trpc2	1315F1	TCATCGGGACCGCCTTYTNTGYGG	50
	1315F2	CCTGTGCGGCATCAACAAYRTNTAYGT	53
	1315R1	TGGGGGTGGGGATGAYRTTRAANGG	50
	1315R2	CGGGCATGGTCAGGCCYTCNCKRAA	53
	1215F	AAAGACGACCCCCAGATCAT	55
	1215R	GCAAACCTTCCACTCCACGTC	55

5.7 LITERATURE CITED

- Adler, E., M. A. Hoon, K. L. Mueller, J. Chandrashekar, N. J. Ryba, and C. S. Zuker. 2000. A novel family of mammalian taste receptors. *Cell* **100**:693-702.
- Alder, M. N., I. B. Rogozin, L. M. Iyer, G. V. Glazko, M. D. Cooper, and Z. Pancer. 2005. Diversity and function of adaptive immune receptors in a jawless vertebrate. *Science* **310**:1970-1973.
- Arendt, D. 2003. Evolution of eyes and photoreceptor cell types. *Int J Dev Biol* **47**:563-571.
- Arendt, D., K. Tessmar-Raible, H. Snyman, A. W. Dorresteijn, and J. Wittbrodt. 2004. Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. *Science* **306**:869-871.
- Bartl, S., M. A. Baish, M. F. Flajnik, and Y. Ohta. 1997. Identification of class I genes in cartilaginous fish, the most ancient group of vertebrates displaying an adaptive immune response. *J Immunol* **159**:6097-6104.
- Baxi, K. N., K. M. Dorries, and H. L. Eisthen. 2006. Is the vomeronasal system really specialized for detecting pheromones? *Trends Neurosci* **29**:1-7.
- Berghard, A., and L. Dryer. 1998. A novel family of ancient vertebrate odorant receptors. *J Neurobiol* **37**:383-392.
- Bertmar, G. 1981. Evolution of vomeronasal organs in vertebrates. *Evolution* **35**:359-366.
- Cao, Y., B. C. Oh, and L. Stryer. 1998. Cloning and localization of two multigene receptor families in goldfish olfactory epithelium. *Proc Natl Acad Sci U S A* **95**:11987-11992.
- Date-Ito, A., H. Ohara, M. Ichikawa, Y. Mori, and K. Hagino-Yamagishi. 2008. Xenopus VIR Vomeronasal Receptor Family Is Expressed in the Main Olfactory System. *Chem Senses*.
- Dukes, J. P., R. Deaville, M. W. Bruford, A. F. Youngson, and W. C. Jordan. 2004. Odorant receptor gene expression changes during the parr-smolt transformation in Atlantic salmon. *Mol Ecol* **13**:2851-2857.
- Dulac, C., and R. Axel. 1995. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83**:195-206.
- Dulka, J. G. 1993. Sex pheromone systems in goldfish: comparisons to vomeronasal systems in tetrapods. *Brain Behav Evol* **42**:265-280.
- Eisthen, H. L. 1992. Phylogeny of the vomeronasal system and of receptor cell types in the olfactory and vomeronasal epithelia of vertebrates. *Microsc Res Tech* **23**:1-21.
- Eisthen, H. L. 1997. Evolution of vertebrate olfactory systems. *Brain Behav Evol* **50**:222-233.
- Eisthen, H. L. 2000. Presence of the vomeronasal system in aquatic salamanders. *Philos Trans R Soc Lond B Biol Sci* **355**:1209-1213.
- Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**:783-791.
- Freitag, J., A. Beck, G. Ludwig, L. von Buchholtz, and H. Breer. 1999. On the origin of the olfactory receptor family: receptor genes of the jawless fish (*Lampetra fluviatilis*). *Gene* **226**:165-174.

- Galant, R., J. B. Skeath, S. Paddock, D. L. Lewis, and S. B. Carroll. 1998. Expression pattern of a butterfly achaete-scute homolog reveals the homology of butterfly wing scales and insect sensory bristles. *Curr Biol* **8**:807-813.
- Gehring, W. J. 1996. The master control gene for morphogenesis and evolution of the eye. *Genes Cells* **1**:11-15.
- Go, Y., Y. Satta, O. Takenaka, and N. Takahata. 2005. Lineage-specific loss of function of bitter taste receptor genes in humans and nonhuman primates. *Genetics* **170**:313-326.
- Grus, W. E., and J. Zhang. 2006. Origin and evolution of the vertebrate vomeronasal system viewed through system-specific genes. *Bioessays* **28**:709-718.
- Hagino-Yamagishi, K., K. Moriya, H. Kubo, Y. Wakabayashi, N. Isobe, S. Saito, M. Ichikawa, and K. Yazaki. 2004. Expression of vomeronasal receptor genes in *Xenopus laevis*. *J Comp Neurol* **472**:246-256.
- Hansen, A., K. T. Anderson, and T. E. Finger. 2004. Differential distribution of olfactory receptor neurons in goldfish: structural and molecular correlates. *J Comp Neurol* **477**:347-359.
- Hansen, A., S. H. Rolen, K. Anderson, Y. Morita, J. Caprio, and T. E. Finger. 2003. Correlation between olfactory receptor cell type and function in the channel catfish. *J Neurosci* **23**:9328-9339.
- Herrada, G., and C. Dulac. 1997. A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. *Cell* **90**:763-773.
- Hoon, M. A., E. Adler, J. Lindemeier, J. F. Battey, N. J. Ryba, and C. S. Zuker. 1999. Putative mammalian taste receptors: a class of taste-specific GPCRs with distinct topographic selectivity. *Cell* **96**:541-551.
- Iwasaki, S. 2002. Evolution of the structure and function of the vertebrate tongue. *J Anat* **201**:1-13.
- Jermakowicz, W. J., 3rd, D. A. Dorsey, A. L. Brown, K. Wojciechowski, C. L. Giscombe, B. M. Graves, C. H. Summers, and G. R. Ten Eyck. 2004. Development of the nasal chemosensory organs in two terrestrial anurans: The directly developing frog, *Eleutherodactylus coqui* (Anura: Leptodactylidae), and the metamorphosing toad, *Bufo americanus* (Anura: Bufonidae). *J Morphol* **261**:225-248.
- Kelliher, K. R. 2007. The combined role of the main olfactory and vomeronasal systems in social communication in mammals. *Horm Behav*.
- Kimchi, T., J. Xu, and C. Dulac. 2007. A functional circuit underlying male sexual behaviour in the female mouse brain. *Nature* **448**:1009-1014.
- Laframboise, A. J., X. Ren, S. Chang, R. Dubuc, and B. S. Zielinski. 2007. Olfactory sensory neurons in the sea lamprey display polymorphisms. *Neurosci Lett* **414**:277-281.
- Leybold, B. G., C. R. Yu, T. Leinders-Zufall, M. M. Kim, F. Zufall, and R. Axel. 2002. Altered sexual and social behaviors in *trp2* mutant mice. *Proc Natl Acad Sci U S A* **99**:6376-6381.
- Li, W., A. P. Scott, M. J. Siefkes, H. Yan, Q. Liu, S. S. Yun, and D. A. Gage. 2002. Bile Acid secreted by male sea lamprey that acts as a sex pheromone. *Science* **296**:138-141.

- Li, W., P. W. Sorensen, and D. D. Gallaher. 1995. The olfactory system of migratory adult sea lamprey (*Petromyzon marinus*) is specifically and acutely sensitive to unique bile acids released by conspecific larvae. *J Gen Physiol* **105**:569-587.
- Liman, E. R., D. P. Corey, and C. Dulac. 1999. TRP2: a candidate transduction channel for mammalian pheromone sensory signaling. *Proc Natl Acad Sci U S A* **96**:5791-5796.
- Liman, E. R., and H. Innan. 2003. Relaxed selective pressure on an essential component of pheromone transduction in primate evolution. *Proc Natl Acad Sci U S A* **100**:3328-3332.
- Matsunami, H., and L. B. Buck. 1997. A multigene family encoding a diverse array of putative pheromone receptors in mammals. *Cell* **90**:775-784.
- Mombaerts, P. 2004. Genes and ligands for odorant, vomeronasal and taste receptors. *Nat Rev Neurosci* **5**:263-278.
- Naito, T., Y. Saito, J. Yamamoto, Y. Nozaki, K. Tomura, M. Hazama, S. Nakanishi, and S. Brenner. 1998. Putative pheromone receptors related to the Ca²⁺-sensing receptor in Fugu. *Proc Natl Acad Sci U S A* **95**:5178-5181.
- Nei, M., and S. Kumar. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- Oike, H., T. Nagai, A. Furuyama, S. Okada, Y. Aihara, Y. Ishimaru, T. Marui, I. Matsumoto, T. Misaka, and K. Abe. 2007. Characterization of ligands for fish taste receptors. *J Neurosci* **27**:5584-5592.
- Okabe, M., and A. Graham. 2004. The origin of the parathyroid gland. *Proc Natl Acad Sci U S A* **101**:17716-17719.
- Pancer, Z., N. R. Saha, J. Kasamatsu, T. Suzuki, C. T. Amemiya, M. Kasahara, and M. D. Cooper. 2005. Variable lymphocyte receptors in hagfish. *Proc Natl Acad Sci U S A* **102**:9224-9229.
- Perez, C. A., L. Huang, M. Rong, J. A. Kozak, A. K. Preuss, H. Zhang, M. Max, and R. F. Margolskee. 2002. A transient receptor potential channel expressed in taste receptor cells. *Nat Neurosci* **5**:1169-1176.
- Pfister, P., J. Randall, J. I. Montoya-Burgos, and I. Rodriguez. 2007. Divergent Evolution among Teleost V1r Receptor Genes. *PLoS ONE* **2**:e379.
- Pfister, P., and I. Rodriguez. 2005. Olfactory expression of a single and highly variable V1r pheromone receptor-like gene in fish species. *Proc Natl Acad Sci U S A* **102**:5489-5494.
- Prachumwat, A., and W. H. Li. 2008. Gene number expansion and contraction in vertebrate genomes with respect to invertebrate genomes. *Genome Res* **18**:221-232.
- Restrepo, D., J. Arellano, A. M. Oliva, M. L. Schaefer, and W. Lin. 2004. Emerging views on the distinct but related roles of the main and accessory olfactory systems in responsiveness to chemosensory signals in mice. *Horm Behav* **46**:247-256.
- Rose, T. M., J. G. Henikoff, and S. Henikoff. 2003. CODEHOP (Consensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. *Nucleic Acids Res* **31**:3763-3766.
- Rose, T. M., E. R. Schultz, J. G. Henikoff, S. Pietrokovski, C. M. McCallum, and S. Henikoff. 1998. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res* **26**:1628-1635.

- Ryba, N. J., and R. Tirindelli. 1997. A new multigene family of putative pheromone receptors. *Neuron* **19**:371-379.
- Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**:406-425.
- Saraiva, L. R., and S. I. Korsching. 2007. A novel olfactory receptor gene family in teleost fish. *Genome Res* **17**:1448-1457.
- Sato, Y., N. Miyasaka, and Y. Yoshihara. 2005. Mutually exclusive glomerular innervation by two distinct types of olfactory sensory neurons revealed in transgenic zebrafish. *J Neurosci* **25**:4889-4897.
- Satoh, G. 2005. Characterization of novel GPCR gene coding locus in amphioxus genome: gene structure, expression, and phylogenetic analysis with implications for its involvement in chemoreception. *Genesis* **41**:47-57.
- Scalia, F., and S. S. Winans. 1975. The differential projections of the olfactory bulb and accessory olfactory bulb in mammals. *J Comp Neurol* **161**:31-55.
- Serb, J. M., and T. H. Oakley. 2005. Hierarchical phylogenetics as a quantitative analytical framework for evolutionary developmental biology. *Bioessays* **27**:1158-1166.
- Shi, P., J. P. Bielawski, H. Yang, and Y. P. Zhang. 2005. Adaptive diversification of vomeronasal receptor 1 genes in rodents. *J Mol Evol* **60**:566-576.
- Shi, P., and J. Zhang. 2006. Contrasting modes of evolution between vertebrate sweet/umami receptor genes and bitter receptor genes. *Mol Biol Evol* **23**:292-300.
- Shi, P., and J. Zhang. 2007. Comparative genomic analysis identifies an evolutionary shift of vomeronasal receptor gene repertoires in the vertebrate transition from water to land. *Genome Res* **17**:166-174.
- Siefkes, M. J., and W. Li. 2004. Electrophysiological evidence for detection and discrimination of pheromonal bile acids by the olfactory epithelium of female sea lampreys (*Petromyzon marinus*). *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* **190**:193-199.
- Spehr, M., J. Spehr, K. Ukhanov, K. R. Kelliher, T. Leinders-Zufall, and F. Zufall. 2006. Parallel processing of social signals by the mammalian main and accessory olfactory systems. *Cell Mol Life Sci* **63**:1476-1484.
- Stowers, L., T. E. Holy, M. Meister, C. Dulac, and G. Koentges. 2002. Loss of sex discrimination and male-male aggression in mice deficient for TRP2. *Science* **295**:1493-1500.
- Takigami, S., Y. Mori, Y. Tanioka, and M. Ichikawa. 2004. Morphological evidence for two types of Mammalian vomeronasal system. *Chem Senses* **29**:301-310.
- Taniguchi, K., and K. Taniguchi. 2008. Embryonic and postnatal differentiation of olfactory epithelium and vomeronasal organ in the Syrian hamster. *J Vet Med Sci* **70**:57-64.
- Taniguchi, K., Y. Toshima, T. R. Saito, and K. Taniguchi. 1996. Development of the olfactory epithelium and vomeronasal organ in the Japanese reddish frog, *Rana japonica*. *J Vet Med Sci* **58**:7-15.
- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**:4876-4882.

- Venkatesh, B., E. F. Kirkness, Y. H. Loh, A. L. Halpern, A. P. Lee, J. Johnson, N. Dandona, L. D. Viswanathan, A. Tay, J. C. Venter, R. L. Strausberg, and S. Brenner. 2007. Survey sequencing and comparative analysis of the elephant shark (*Callorhynchus milii*) genome. *PLoS Biol* **5**:e101.
- Wray, G. A., and E. Abouheif. 1998. When is homology not homology? *Curr Opin Genet Dev* **8**:675-680.
- Young, J. M., and B. J. Trask. 2007. V2R gene families degenerated in primates, dog and cow, but expanded in opossum. *Trends Genet*.
- Zhang, J., and D. M. Webb. 2003. Evolutionary deterioration of the vomeronasal pheromone transduction pathway in catarrhine primates. *Proc Natl Acad Sci U S A* **100**:8337-8341.

CONCLUSION

Using the genes specific to the vomeronasal system (VNS), I addressed three questions concerning the evolution of the VNS that were previously unanswerable without both knowledge of system-specific genes and full genome sequences. First, I investigate how the V1R vomeronasal receptor family evolves in mammals and how the evolution of V1Rs reflects known variation in VNS morphological components. Next, I addressed how the evolutionary patterns of VNS receptors compare to the evolutionary patterns of main olfactory system and evaluated the results in terms of the two systems' potential functional differences. Finally, I determined when during vertebrate evolution the VNS evolved.

In Chapter 1, I show that species-specificity of V1R vomeronasal repertoires in mouse and rat evolved via gene-sorting evolution following the birth-and-death model of gene family evolution. As such, frequent gene gain and gene loss has characterized this gene family in these two species resulting in a very low frequency of V1R orthologous mouse-rat pairs. Additionally, gene loss led to two mouse-specific and two rat-specific V1R subfamilies.

In Chapter 2, I further explore the evolution of the V1R family in a wider range of mammalian taxa with different levels of vomeronasal organ complexity. Despite what had been shown previously for humans, mouse, and rat, we showed that not all mammalian taxa have a large V1R repertoire size. Instead, we found that the V1R gene

family has the largest functional size ratio (from 187 in mouse to 8 in dog) of any mammalian gene family. Furthermore, we find that repetitive element activity likely led to the larger gene family size in rodents compared to dog and cow. Finally, we find that a representative marsupial, the opossum, has had independent expansions in the same V1R family. We find that the functional V1R repertoire size across all these mammalian taxa is positively correlated to the complexity of the vomeronasal organ.

In Chapter 3, we examine the evolution of the V1Rs in the final of the three mammalian lineages, the monotremes. In congruence with our previously identified positive correlation between vomeronasal organ complexity and V1R repertoire, we find that the platypus, which has the most complex type of vomeronasal organ, has the largest V1R repertoire size so far identified. This larger V1R repertoire size is larger than the odorant receptor (OR) family in platypus, despite the OR family being the largest mammalian gene family in most species. Thus, the platypus clearly relies heavily on vomeronasal-mediated olfaction despite its semi-aquatic lifestyle.

In Chapter 4, we show that the evolutionary patterns in the vomeronasal receptors (V1Rs and V2Rs) are clearly and quantitatively distinct from the evolutionary patterns of the main olfactory receptors (ORs and TAARs). The patterns revealed by both types of receptors support the differential tuning hypothesis, which suggests that main olfactory receptors are broadly tuned generalists while vomeronasal receptors are narrowly tuned specialists. We determined that ligand-receptor pairings are likely conserved in mouse-rat orthologous pairs, while function changes soon after duplication in closely related paralogs.

In Chapter 5, we use the VNS-specific genes to determine that the VNS was present in the common ancestor of vertebrates. Both V1Rs and Trpc2 were found in the genome of the sea lamprey, and all three VNS-specific genes, V1Rs, V2Rs, and Trpc2, were found in the genome of the elephant shark. The sea lamprey VNS-specific genes are expressed in the olfactory epithelium, suggesting the ancestral chemosensory function for this system.

APPENDIX 1

ORTHOLOGOUS MOUSE-RAT NASAL CHEMOSENSORY RECEPTOR PAIRS

A1.1 Mouse and rat orthologous V1Rs.

Mouse gene names and some rat gene names come from Shi, P., et al. (2005) Adaptive diversification of vomeronasal receptor 1 genes in rodents. *J Mol Evol* 60, 566-576 and the remaining rat gene names come from Shi, P., and Zhang, J. (2007) Comparative genomic analysis identifies an evolutionary shift of vomeronasal receptor gene repertoires in the vertebrate transition from water to land. *Genome Res* 17, 166-174.

MOUSE	RAT
mV1rC28	rv1rC18
mV1rC30	rv1rC10
mV1rC5	rv1rC22
mV1rC6	rv1rC23
mV1rD6	rv1rD1
mV1rE1	rv1rE24
mV1rE13	rv1rE8
mV1rE2	rnew6
mV1rE3	rv1rE22
mV1rE4	rv1rE23
mV1rF2	rv1rF5
mV1rF3	rv1rF4
mV1rG1	rv1rG3
mV1rG11	rv1rG13
mV1rG6	rv1rG4
mV1rJ3	rv1rJ1
mV1rK1	rv1rK1
mV1rL1	rv1rL1

A1.2 Mouse and rat orthologous V2Rs.

Gene names come from Yang, H., et al. (2005) Composition and evolution of the V2r vomeronasal receptor gene repertoire in mice and rats. *Genomics* 86, 306-315.

MOUSE	RAT
V2r49	V2r133
V2r55	V2r101
V2r63	V2r148
v2r93	V2r152

A1.3 Orthologous ORs for mouse and rat.

Mouse receptor names are from Young, J.M., et al. (2002) Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Hum Mol Genet* 11, 535-546 and Zhang, X., and Firestein, S. (2002) The olfactory receptor gene superfamily of the mouse. *Nat Neurosci* 5, 124-133. Rat gene names are from Rat Genome Sequencing Consortium (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428, 493-521.

Mouse	Rat
mouse1006	Olr450
mouse1008	Olr490
mouse1009	Olr453
mouse1010	Olr454
mouse1012	Olr455
mouse1013	Olr456
mouse1015	Olr458
mouse1016	Olr459
mouse1018	Olr461
mouse1019	Olr462
mouse1020	Olr463
mouse1022	Olr465
mouse1023	Olr466
mouse1026	Olr469
mouse103	Olr1734
mouse1030	Olr471
mouse1031	Olr472
mouse1032	Olr473
mouse1037	Olr476
mouse1038	Olr477
mouse1042	Olr480
mouse1043	Olr481
mouse1044	Olr482
mouse1045	Olr484
mouse1046	Olr485
mouse1047	Olr486
mouse1054	Olr491
mouse1056	Olr493
mouse1058	Olr496
mouse1065	Olr499
mouse1079	Olr507
mouse108	Olr1714
mouse1086	Olr510
mouse1089	Olr513
mouse1090	Olr514
mouse1093	Olr515
mouse1099	Olr529
mouse1101	Olr533
mouse1104	Olr536
mouse1105	Olr537
mouse1106	Olr539
mouse1109	Olr541

mouse1111	Olr544
mouse1112	Olr554
mouse1120	Olr558
mouse1121	Olr559
mouse1122	Olr560
mouse1123	Olr561
mouse1125	Olr562
mouse1126	Olr566
mouse1132	Olr576
mouse1134	Olr577
mouse1135	Olr578
mouse1136	Olr583
mouse1137	Olr581
mouse1138	Olr584
mouse1140	Olr586
mouse1141	Olr587
mouse1143	Olr590
mouse1145	Olr592
mouse1148	Olr595
mouse1151	Olr596
mouse1152	Olr597
mouse1153	Olr598
mouse1156	Olr601
mouse1157	Olr602
mouse1163	Olr607
mouse1164	Olr608
mouse1165	Olr609
mouse1170	Olr610
mouse1178	Olr619
mouse1180	Olr630
mouse1189	Olr646
mouse1195	Olr651
mouse1199	Olr657
mouse12	Olr1353
mouse1208	Olr661
mouse1209	Olr662
mouse1211	Olr663
mouse1212	Olr664
mouse1215	Olr670
mouse1216	Olr671
mouse1217	Olr672
mouse1218	Olr673
mouse122	Olr1696
mouse1220	Olr677
mouse1221	Olr678
mouse1222	Olr681
mouse1225	Olr686
mouse1226	Olr687
mouse1228	Olr689
mouse1229	Olr690
mouse123	Olr1695
mouse1230	Olr693

mouse1231	Olr694
mouse1232	Olr695
mouse1233	Olr696
mouse1234	Olr697
mouse1238	Olr703
mouse1239	Olr701
mouse124	Olr1694
mouse1241	Olr704
mouse1242	Olr705
mouse1243	Olr684
mouse1245	Olr707
mouse1246	Olr709
mouse1247	Olr710
mouse125	Olr1693
mouse1250	Olr711
mouse1251	Olr712
mouse1254	Olr718
mouse1256	Olr720
mouse1257	Olr722
mouse1260	Olr725
mouse1262	Olr727
mouse1263	Olr728
mouse1264	Olr729
mouse1270	Olr741
mouse1275	Olr747
mouse1276	Olr748
mouse1277	Olr750
mouse1279	Olr752
mouse1280	Olr753
mouse1284	Olr757
mouse1288	Olr758
mouse129	Olr1687
mouse1290	Olr760
mouse1294	Olr766
mouse1295	Olr767
mouse1299	Olr770
mouse13	Olr818
mouse1301	Olr772
mouse1302	Olr773
mouse1305	Olr778
mouse1308	Olr781
mouse1309	Olr782
mouse1310	Olr783
mouse1311	Olr784
mouse1312	Olr785
mouse132	Olr1684
mouse1321	Olr1765
mouse1322	Olr1766
mouse1323	Olr1767
mouse1324	Olr1751
mouse1335	Olr862
mouse1336	Olr8

mouse1339	Olr859
mouse1341	Olr858
mouse1342	Olr857
mouse1346	Olr7
mouse1347	Olr5
mouse1348	Olr4
mouse1349	Olr3
mouse1351	Olr1076
mouse1352	Olr1077
mouse1353	Olr1078
mouse1356	Olr1086
mouse1357	Olr1087
mouse1361	Olr1657
mouse1364	Olr1658
mouse1368	Olr1660
mouse137	Olr1671
mouse1370	Olr1664
mouse1377	Olr1405
mouse1378	Olr1404
mouse1383	Olr1398
mouse1384	Olr1397
mouse1387	Olr1393
mouse1388	Olr1392
mouse1389	Olr1391
mouse1391	Olr1389
mouse140	Olr731
mouse1402	Olr390
mouse1403	Olr1576
mouse1404	Olr1583
mouse1406	Olr1581
mouse1408	Olr1579
mouse1411	Olr1350
mouse1412	Olr1349
mouse1414	Olr1345
mouse1416	Olr1343
mouse1420	Olr319
mouse1423	Olr321
mouse1426	Olr323
mouse1427	Olr324
mouse1428	Olr325
mouse1431	Olr329
mouse1440	Olr336
mouse1441	Olr337
mouse1443	Olr338
mouse1444	Olr339
mouse1445	Olr340
mouse1448	Olr343
mouse1449	Olr344
mouse145	Olr1201
mouse1451	Olr346
mouse1454	Olr348
mouse1459	Olr360

mouse1461	Olr361
mouse1462	Olr352
mouse1469	Olr363
mouse148	Olr1308
mouse1489	Olr371
mouse149	Olr1315
mouse1490	Olr372
mouse1494	Olr374
mouse1495	Olr375
mouse1496	Olr376
mouse1497	Olr377
mouse1499	Olr382
mouse15	Olr1356
mouse1500	Olr379
mouse151	Olr1195
mouse1510	Olr1642
mouse1511	Olr1641
mouse1512	Olr1640
mouse1514	Olr1638
mouse152	Olr588
mouse155	Olr840
mouse156	Olr838
mouse157	Olr839
mouse159	Olr836
mouse161	Olr1373
mouse164	Olr1570
mouse166	Olr1569
mouse167	Olr1568
mouse174	Olr1564
mouse178	Olr1560
mouse186	Olr1551
mouse195	Olr1542
mouse196	Olr1535
mouse20	Olr1466
mouse204	Olr1532
mouse205	Olr1530
mouse206	Olr1529
mouse211	Olr823
mouse214	Olr830
mouse222	Olr1462
mouse223	Olr1463
mouse24	Olr1121
mouse25	Olr1228
mouse266	Olr392
mouse267	Olr854
mouse270	Olr853
mouse272	Olr851
mouse273	Olr847
mouse275	Olr841
mouse281	Olr1108
mouse282	Olr1107
mouse283	Olr1106

mouse284	Olr1105
mouse285	Olr1104
mouse299	Olr230
mouse3	Olr411
mouse30	Olr1413
mouse304	Olr160
mouse307	Olr140
mouse308	Olr1310
mouse309	Olr120
mouse31	Olr1606
mouse310	Olr200
mouse317	Olr1433
mouse32	Olr737
mouse323	Olr1432
mouse324	Olr1431
mouse328	Olr1425
mouse33	Olr62
mouse332	Olr1415
mouse338	Olr399
mouse341	Olr401
mouse342	Olr402
mouse344	Olr403
mouse345	Olr404
mouse348	Olr409
mouse350	Olr413
mouse351	Olr414
mouse352	Olr415
mouse353	Olr416
mouse354	Olr417
mouse355	Olr418
mouse356	Olr419
mouse357	Olr420
mouse358	Olr421
mouse361	Olr424
mouse362	Olr425
mouse365	Olr428
mouse366	Olr429
mouse368	Olr434
mouse370	Olr1667
mouse371	Olr1666
mouse376	Olr1467
mouse378	Olr1470
mouse38	Olr807
mouse382	Olr1485
mouse385	Olr1486
mouse39	Olr1186
mouse392	Olr1498
mouse393	Olr1499
mouse398	Olr1507
mouse399	Olr1509
mouse401	Olr1511
mouse41	Olr226

mouse414	Olr1601
mouse419	Olr1597
mouse42	Olr1662
mouse420	Olr1596
mouse421	Olr1595
mouse424	Olr1593
mouse429	Olr1589
mouse435	Olr819
mouse438	Olr816
mouse44	Olr1302
mouse447	Olr813
mouse448	Olr812
mouse449	Olr811
mouse450	Olr810
mouse452	Olr808
mouse453	Olr806
mouse458	Olr801
mouse460	Olr799
mouse462	Olr1521
mouse464	Olr1523
mouse466	Olr1652
mouse469	Olr237
mouse474	Olr241
mouse478	Olr244
mouse479	Olr245
mouse48	Olr721
mouse481	Olr246
mouse483	Olr250
mouse49	Olr1646
mouse490	Olr259
mouse491	Olr260
mouse493	Olr262
mouse494	Olr270
mouse497	Olr263
mouse5	Olr6
mouse50	Olr410
mouse506	Olr268
mouse507	Olr272
mouse51	Olr1407
mouse510	Olr276
mouse514	Olr285
mouse516	Olr281
mouse517	Olr283
mouse518	Olr282
mouse52	Olr483
mouse522	Olr2861
mouse523	Olr287
mouse524	Olr288
mouse530	Olr297
mouse531	Olr298
mouse536	Olr303
mouse539	Olr312

mouse544	Olr39
mouse545	Olr40
mouse547	Olr41
mouse550	Olr43
mouse551	Olr44
mouse552	Olr45
mouse553	Olr47
mouse554	Olr48
mouse558	Olr63
mouse56	Olr1386
mouse561	Olr57
mouse564	Olr56
mouse568	Olr51
mouse569	Olr50
mouse57	Olr1075
mouse571	Olr49
mouse577	Olr69
mouse578	Olr70
mouse584	Olr74
mouse589	Olr80
mouse591	Olr82
mouse593	Olr84
mouse594	Olr85
mouse599	Olr91
mouse60	Olr295
mouse600	Olr92
mouse601	Olr93
mouse604	Olr87
mouse605	Olr95
mouse606	Olr96
mouse609	Olr104
mouse61	Olr311
mouse610	Olr105
mouse611	Olr106
mouse614	Olr107
mouse615	Olr108
mouse616	Olr109
mouse617	Olr110
mouse618	Olr111
mouse619	Olr112
mouse62	Olr855
mouse620	Olr113
mouse623	Olr119
mouse624	Olr120
mouse627	Olr122
mouse629	Olr124
mouse630	Olr125
mouse633	Olr135
mouse639	Olr141
mouse64	Olr132
mouse640	Olr142
mouse641	Olr143

mouse642	Olr144
mouse646	Olr148
mouse648	Olr149
mouse649	Olr150
mouse65	Olr131
mouse651	Olr152
mouse652	Olr153
mouse653	Olr154
mouse654	Olr155
mouse655	Olr156
mouse656	Olr157
mouse659	Olr160
mouse66	Olr130
mouse661	Olr161
mouse665	Olr171
mouse667	Olr175
mouse668	Olr176
mouse67	Olr128
mouse670	Olr178
mouse676	Olr185
mouse677	Olr188
mouse68	Olr127
mouse683	Olr192
mouse684	Olr193
mouse686	Olr197
mouse690	Olr201
mouse691	Olr202
mouse695	Olr203
mouse699	Olr209
mouse70	Olr833
mouse700	Olr208
mouse701	Olr210
mouse704	Olr215
mouse705	Olr217
mouse706	Olr218
mouse71	Olr834
mouse710	Olr222
mouse711	Olr224
mouse713	Olr227
mouse715	Olr233
mouse716	Olr234
mouse720	Olr1607
mouse722	Olr1624
mouse727	Olr1627
mouse728	Olr1629
mouse729	Olr1630
mouse730	Olr1608
mouse731	Olr1609
mouse732	Olr1610
mouse734	Olr1612
mouse736	Olr1614
mouse738	Olr1615

mouse739	Olr1616
mouse74	Olr604
mouse744	Olr1620
mouse750	Olr1637
mouse76	Olr326
mouse768	Olr879
mouse769	Olr880
mouse77	Olr1163
mouse770	Olr881
mouse774	Olr906
mouse78	Olr59
mouse780	Olr1006
mouse786	Olr964
mouse790	Olr1014
mouse791	Olr892
mouse794	Olr889
mouse796	Olr1016
mouse801	Olr1020
mouse808	Olr947
mouse811	Olr1059
mouse812	Olr1055
mouse813	Olr932
mouse818	Olr1060
mouse821	Olr1064
mouse826	Olr1071
mouse827	Olr1072
mouse828	Olr1122
mouse829	Olr1118
mouse834	Olr1125
mouse835	Olr1128
mouse843	Olr1138
mouse845	Olr1126
mouse847	Olr1147
mouse850	Olr1144
mouse851	Olr1148
mouse855	Olr1156
mouse867	Olr1172
mouse868	Olr1179
mouse874	Olr1196
mouse876	Olr1199
mouse877	Olr1200
mouse878	Olr1202
mouse881	Olr1204
mouse888	Olr1213
mouse889	Olr1214
mouse891	Olr1219
mouse893	Olr1222
mouse894	Olr1223
mouse895	Olr1225
mouse898	Olr1230
mouse90	Olr1750
mouse901	Olr1236

mouse904	Olr1240
mouse905	Olr1241
mouse906	Olr1242
mouse907	Olr1243
mouse908	Olr1244
mouse909	Olr1245
mouse91	Olr1749
mouse913	Olr1250
mouse914	Olr1251
mouse915	Olr1252
mouse916	Olr1253
mouse917	Olr1254
mouse92	Olr1746
mouse920	Olr1259
mouse922	Olr1264
mouse923	Olr1265
mouse93	Olr1748
mouse930	Olr1271
mouse934	Olr1273
mouse937	Olr1278
mouse938	Olr1279
mouse943	Olr1293
mouse945	Olr1295
mouse948	Olr1294
mouse95	Olr1744
mouse958	Olr1306
mouse959	Olr1307
mouse96	Olr1743
mouse960	Olr1309
mouse961	Olr1311
mouse963	Olr1313
mouse967	Olr1321
mouse97	Olr1742
mouse974	Olr1329
mouse975	Olr1330
mouse976	Olr1331
mouse979	Olr1334
mouse980	Olr1335
mouse982	Olr1337
mouse983	Olr1338
mouse984	Olr1339
mouse986	Olr1341
mouse992	Olr439
mouse993	Olr440
mouse994	Olr441
mouse995	Olr442
mouse998	Olr446

A1.4 Mouse and rat orthologous TAARs.

Gene names correspond to names from Lindemann, L., et al. (2005) Trace amine-associated receptors form structurally and functionally distinct subfamilies of novel G protein-coupled receptors. *Genomics* 85, 372-385.

MOUSE	RAT
mTAAR1	rTAAR1
mTAAR2	rTAAR2
mTAAR3	rTAAR3
mTAAR4	rTAAR4
mTAAR5	rTAAR5
mTAAR6	rTAAR6
mTAAR9	rTAAR9