More Functional V1R Genes Occur in Nest-Living and Nocturnal Terricolous Mammals

Guodong Wang¹, Peng Shi*,¹, Zhouhai Zhu², and Ya-ping Zhang*,¹,²

¹State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China
²Laboratory for Conservation and Utilization of Bio-resources, Yunnan University, Kunming, China

*Corresponding author: E-mail: zhangyp1@263.net.cn; shipengsir@gmail.com.

Accepted: 5 May 2010

Abstract

Size of the vomeronasal type 1 receptor (V1R) gene repertoire may be a good indicator for examining the relationship between animal genomes and their environmental niche specialization, especially the relationship between ecological factors and the molecular evolutionary history of the sensory system. Recently, Young et al. (Young JM, Massa HF, Hsu L, Trask BJ. 2009. Extreme variability among mammalian V1R gene families. Genome Res.) concluded that no single ecological factor could explain the extreme variability of the V1R gene repertoire in mammalian genomes. In contrast, we found a significant positive correlation between the size and percentage of intact V1R genes in 32 species that represent the phylogenetic diversity of terricolous mammals and two ecological factors: spatial activity and rhythm activity. Nest-living species possessed a greater number of intact V1R genes than open-living species, and nocturnal terricolous mammals tended to possess more intact V1R genes than diurnal species. Moreover, our analysis reveals that the evolutionary mechanisms underlying these observations likely resulted from the rapid gene birth and accelerated amino acid substitutions in nest-living and nocturnal mammals, likely a functional requirement for exploiting narrow, dark environments. Taken together, these results reveal how adaptation to divergent circadian rhythms and spatial activity were manifested at the genomic scale. Size of the V1R gene family might have indicated how this gene family adapts to ecological factors.

Key words: V1R repertoire, spatial activity, rhythm activity, adaptive selection.

Pheromones are a group of chemical signals that trigger intraspecific behavioral responses, such as social and reproductive behaviors (Halpern 1987; Keverne 1999, 2002; Halpern and Martinez-Marcos 2003). In mammals, they play multiple critical roles in daily life, including the recognition of individuals, mating, and territoriality (Prasad and Reed 1999). The vomeronasal organ (VNO) appears to be specialized in the detection of pheromones, although the main olfactory epithelium can also detect some of them (Boehm et al. 2005; Yoon et al. 2005). At least three gene families (vomeronasal type 1 receptor [V1R], vomeronasal type 2 receptor, and formyl peptide receptors) are expressed in the VNO, and the encoded chemosensory receptors directly interact with the external environment (Dulac and Axel 1995; Matsunami and Buck 1997; Rivière et al. 2009). The V1R gene family provides an excellent opportunity to study the molecular basis of adaptation in mammalian behaviors, lifestyles, and environments because it exhibits the greatest among-species variation in gene family size of all the mammalian gene families (Grus et al. 2005, 2007; Young et al. 2005, 2009; Shi and Zhang 2007, 2009). This high level of variation may be functionally linked to the adaptation of species to their specific environments (Shi and Zhang 2009). Uniquely, the size of the V1R gene family is positively correlated with the morphological complexity of the mammalian VNO (Grus et al. 2005, 2007). This suggests that variation in functional V1R gene numbers reflects changes in VNO morphology and function. Indeed, enhancement in the V1R gene repertoires (V1RGRs) during the vertebrate transition from water to land reflects an adaptation to terrestrial life (Shi and Zhang 2007).

Environmentally, mammalian species range from being nocturnal to diurnal and from arboreal to subterranean. The sense of smell is used extensively by some species to establish and maintain communication (Burda et al. 1990; Francescoli 2000). Spatial activity patterns, such as nest- or open-living behavior, likely affect communication channels and chemical signaling. Nest-living rodents depend
more on chemical cues than do terraneous ones because vision is much more limited in dark nests and hollows (Francescoli 2000). Nocturnal animals rely predominantly on olfactory cues to mediate social interactions and sexual communication; these species generally have a highly developed sense of smell (Jaeger and Gergits 1979; Dawley 1984; Mathis 1990; Gillette et al. 2000; Palmer 2004). After estimating the size of the V1RGR from 31 terricolous mammalian genomes, Young et al. (2009) concluded that no single ecological factor explained the extreme variability. However, they did not test specific null hypotheses. Consequently, we conducted an analysis of the V1RGR from 32 mammalian genomes in the context of two ecological factors: spatial activity (nest-living behavior and open-living behavior) and rhythm activity (diurnal and nocturnal) (supplementary table 1, Supplementary Material online). Our two null hypotheses stated that activity periods and terricolous habitats were not correlated with V1R gene family size (Young et al. 2009). A correlation between either of these life styles with the V1RGR would serve to reject either null hypothesis. The alternative hypotheses stated that correlated tendencies would indicate adaptive responses to activity periods, terricolous habitats, or both. Statistical tests of the null hypotheses could reject either null hypotheses. Our results are robust to different statistical tests and provide strong evidence supporting our hypotheses that 1) nest-living mammals generally have more intact V1R genes than do open-living mammals and 2) nocturnal mammals generally have more intact V1R genes than do diurnal mammals. Our work also suggests that the size of the V1R gene family could be a good indicator in studies of the interaction between the evolution of this gene family and ecological factors.

**Nest-Living Terricolous Mammals Have a Greater Numbers of Functional V1R Genes Than Do Open-Living Species**

Are dramatic changes in the structure of the VNO, as well as the size of the V1RGR, correlated with the ecological diversity? To address this question, we first compared the V1RGR in nest-living and open-living species. Nest-living mammals, those living in a narrow space (e.g., burrows), occur where vision and hearing airborne sounds are limited (Burd et al. 1999; Francescoli 2000). Conversely, opening-living (e.g., grasslands and forests) mammals have relatively unlimited vision and audition (Nowak and Walker 1999). We analyzed the relationship between the V1RGR and the spatial activity among 32 terricolous mammals, 31 of which were used by Young et al. (2009). We added and described the V1RGR from the giant panda genome (Li et al. 2010). The 32 terricolous mammals included 2 Australidelphia, 3 Afrotheria, 2 Xenarthra, 17 Euarchontoglires, and 8 Laurasiatheria (fig. 1) (Murphy et al. 2004). These taxa represented the phylogenetic diversity of mammals (supplementary table 1, Supplementary Material online). First, we divided our sample into two groups: nest-living (15 species) and opening-living mammals (17 species). An analysis of similarity (ANOSIM) showed that the V1RGR varied significantly between the two groups ($P < 0.001$; table 1). The number of intact V1R genes averaged 104.5 in the nest-living terricolous mammals, and this was significantly greater than that for the open-living terricolous mammals (25.6 intact V1Rs) ($P < 0.001$; analysis of variance [ANOVA]; table 1). Thus, nest-living species had significantly more V1R genes than open-living mammals. Second, we conducted a linear regression analysis comparing the V1RGR with different spatial activity patterns. The size of the V1RGR was significantly and positively correlated with nest living ($R = 0.663$, $P < 0.001$; fig. 2). These results suggested that nest-living mammals require a greater number of functional V1Rs in order to exploit their relatively narrow environments.

However, these observations could potentially be explained by the random genomic drift hypothesis (Nozawa et al. 2007; Nei et al. 2008), and phylogenetic inertia—closely related species tend to be similar because of shared inheritance rather than independent adaptation (Harvey and Pagel 1991; Fisher and Owens 2004). To distinguish between random drift and functionality, we examined the relationship between spatial activity and the proportion of functional V1R genes. The results of this analysis showed a positive correlation between the percentage of functional genes and nest-living behavior ($R = 0.608$, $P < 0.001$; fig. 2). Whereas the average percentage of functional V1R genes in nest-living mammals was 32.78%, open-living mammals had only 18.14%. Thus, random drift could not explain our observations. Furthermore, to exclude a potential bias resulting from the nonindependence of the phylogenetic relationships, where phylogenetic inertia (closely related species tend to be similar because of shared inheritance rather than through independent adaptation) might have compromised our analyses (Harvey and Pagel 1991; Fisher and Owens 2004). Consequently, we performed a phylogenetically independent contrasts (PICs) analysis (Felsenstein 1985; Pagel 1992). PIC showed the same significant correlations between the size of the intact V1Rs repertoires and spatial activity ($R = 0.555$; $P = 0.001$; table 1) and also between the percentage of intact V1Rs and spatial activity ($R = 0.522$; $P = 0.001$; table 1). This result was maintained after removing the catarrhine primates (chimpanzee, gorilla, orangutan, and macaque) that lack a functional VNO and have lost the vomeronasal signal transduction component (Zhang and Webb 2003). Significant correlations remained between the size of the V1RGR and spatial...
activity ($R = 0.560; P = 0.001$) and the percentage of intact V1Rs and spatial activity ($R = 0.522; P = 0.001$). Thus, phylogenetic inertia was not demonstrated. Our results from both the general pattern and the phylogeny-based studies demonstrated that nest-living terricolous mammals process a greater number of intact V1R genes and also have a higher percentage of functional V1R genes than do open-living terricolous mammals.

### Table 1

<table>
<thead>
<tr>
<th>Ecological Factors</th>
<th>Data Sets</th>
<th>ANOSIM</th>
<th>ANOVA</th>
<th>PIC Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$R^a$</td>
<td>$P$ Value</td>
<td>$F^b$</td>
</tr>
<tr>
<td>Spatial activity</td>
<td>Number of intact V1Rs</td>
<td>0.438</td>
<td>0.000***</td>
<td>23.532</td>
</tr>
<tr>
<td></td>
<td>Proportion of intact V1Rs</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Rhythm activity</td>
<td>Number of intact V1Rs</td>
<td>0.17</td>
<td>0.020*</td>
<td>5.209</td>
</tr>
<tr>
<td></td>
<td>Proportion of intact V1Rs</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Note:** N/A, not applicable. Statistical significant differences are shown by * for $P < 5\%$, ** for $P < 1\%$, and *** for $P < 0.1\%$.

$^a$ Represented ANOSIM $R$ statistic that estimated the difference between within-group dissimilarities and between-group dissimilarities.

$^b$ Represented $F$ statistical value coming from $F$-test, which is used for comparisons of the components of the total deviation.

$^c$ Represented Pearson correlation coefficient.
Rhythm Activity Is Associated with Variation in the Number of Functional V1R Genes

Do nocturnal mammals have a greater number of functional genes than do diurnal species? After dividing species into nest-living versus open-living behavioral groups, the open-living group has 30-fold variation in the size of the intact V1R repertoire, ranging from 3 in humans and gorillas to 89 in the wallaby. Why does such a large difference occur with the same spatial activity period of mammals? Rhythm activity, being either diurnal or nocturnal, might also affect the V1RGR. Twelve of our species are largely nocturnal (Kurumiya and Kawamura 1988), 13 are diurnal, and 4 are cathemeral mammals, which have approximately equal activity periods throughout the 24-h cycle (Tattersall 1979). The number of functional V1R genes was significantly larger in nocturnal mammals (101.8) compared with cathemeral (43.3) and diurnal mammals (36.7) ($P = 0.012$, ANOVA; table 1). The Pearson product-moment correlation ($R = 0.481$, $P = 0.003$) and PIC analyses ($R = 0.423$, $P = 0.009$) showed significant positive correlation between rhythm activity and the V1RGR (table 1), suggesting that nocturnal mammals have a greater number of functional V1R genes. Removal of the four diurnal catharrhines resulted in a significant correlation between the V1RGR and rhythm activity ($R = 0.410$; $P = 0.017$) and the proportion of intact V1Rs and rhythm activity ($R = 0.356$; $P = 0.034$).

This correlation between rhythm activity and V1RGR might be explained by spatial activity alone. In order to distinguish between the effects of rhythm and spatial activities, we controlled one ecological factor and analyzed the other. Among the 17 species in the open-living group, 9 were diurnal, 4 cathemeral, and 4 nocturnal. The Pearson product-moment correlation analysis ($R = 0.678$; $P = 0.001$) showed a significant positive correlation between the V1RGR and rhythm activity. Conversely, rhythm activity was controlled by an analysis of the diurnal group, including four nest-living and nine open-living mammals. A significant positive correlation occurred between the intact V1Rs and nest-living behavior by the Pearson product-moment correlation analysis ($R = 0.855$; $P < 0.001$). Thus, both spatial activity and rhythm activity independently influence the size of the functional V1RGRs in mammals.

To determine whether the above observations can simply be explained due to the effects of low-quality draft genome sequences, we analyzed the 15 species with high-coverage genome sequences ($6 \times$) in which the majority, if not all, of the V1R genes had been identified. The number of genes in the V1RGR averaged 127.2 in the nest-living mammals, which was significantly greater than that of the open-living mammals ($14.6$) ($P < 0.001$). The proportion of V1R genes that were putatively intact in nest-living mammals was 57.8%, and this was significantly larger than that of open-living mammals (17.5%) ($P < 0.001$). Similarly, a significant correlation occurred between the V1RGR and rhythm activity ($R = 0.635$; $P = 0.007$) and between the percentage of intact V1Rs and rhythm activity ($R = 0.499$; $P = 0.035$). Thus, poor genome quality did not cause the correlation.

Both alternative hypotheses are supported by other evidence. Vision and the hearing of airborne sound are relatively unimportant sensory modalities in nest-living species, although vision is used at nest entrances (Burda et al. 1990). Long-lasting odorants are advantageous to mammals living in dark environments (Christiansen 1976); pheromone signals establish and maintain communication (Francescoli 2000). Most nest-living, nocturnal mammals have relatively larger and/or more complex olfactory organs (Takami 2002). For instance, the VNOs are well developed in nocturnal strepsirrhines, yet they are small and extremely variable in platyrhines, and rudimentary in adult diurnal catharrhines (Stephan et al. 1981, 1984; Baron et al. 1983; Dennis et al. 2004). Mammals have both simple and uniform and complex segregated VNOs. The latter type occurs in nocturnal, nest-living mammals only (Takami 2002). Diurnal mammals generally have a well-developed visual system, and this is particularly advanced in Old World monkeys, which have obtained trichromatic color vision via a gene duplication event (Surridge et al. 2003). Conversely, nocturnal mammals always have poorer vision. Indeed, the short-wavelength opsin gene became a pseudogene independently in several nocturnal mammals, such as bush babies, lorises, lemurs, and blind Ehrenberg’s mole rats (David-Gray et al. 2002; Tan et al. 2005). The VNO-mediated nasal sensory system might have been selectively favored in nest-living and nocturnal mammals due to the limited use of a photosensory system.

Conclusions

Size variation in the V1RGR is associated with ecological changes in mammals. Both nest-living and nocturnal terri- culous mammals, versus open-living and diurnal mammals, respectively, have greater numbers of intact V1Rs and higher percentages of intact V1R genes. Conversely, the V1RGR is remarkably conserved among goat, sheep, and cow, which live in similar environments and had similar feeding habits (Ohara et al. 2009). This finding supports the hypothesis that ecological factors affect molecular evolutionary changes in mammalian V1R gene family size. Phylogenetic analysis reveals that both nest-living and nocturnal mammals have frequent, species-specific gains in V1R genes. Species-specific clusters of nest-living mammals have significantly greater number of functional V1R genes than do open-living mammals ($\chi^2 = 24.52$, $P = 7.32 \times 10^{-9}$). Similarly, the number of species-specific clusters in nocturnal mammals is significantly greater than that of diurnal mammals ($\chi^2 = 12.73$, $P = 3.60 \times 10^{-4}$), and species-specific genes are subject
Reinforcing prior results (Shi et al. 2005), V1R gene clusters seem to evolve under positive Darwinian selection in order to discriminate between large and complex pheromonal mixtures. Diversifying selection on newly duplicated, species-specific genes probably enhance the ability to recognize a diverse array of odors encountered in narrow and dark environments. Other ecological traits may also drive variation in the V1RGR and that of the other VNO receptors. Certainly, future ecological investigations of all VNO receptor families will shed light on the genomic evolutionary mechanisms associated with pheromone detection.

**Materials and Methods**

**Identification of V1R Repertoires** The genome assembly of the giant panda (*Ailuropoda melanoleuca*) was downloaded from BGI-Shenzhen (http://sz.genomics.org.cn/). Sequences of the previously described functional V1Rs were retrieved from the literature (Grus et al. 2005, 2007; Shi and Zhang 2007, 2009). V1RGRs were identified following Grus et al. (2005). First, candidate genes were detected from the local databases by conducting a homology search using WU-Blast, with a cutoff $E$ value of $10^{-5}$. Second, the identified putative sequences were Blasted against the non-redundant database of GenBank to ensure V1R gene identity. Open reading frames (ORFs) longer than 270 amino acids that encode protein products and contain the whole putative seven-transmembrane domain were considered to be intact genes. A hit sequence was considered to be a disrupted gene if its disrupted ORF was longer than 200 nucleotides, which usually was incomplete across the 13 (7 transmembrane, 3 extracellular, and 3 intracellular) internal domains.

**Compilation of Ecological Data and Statistical Analysis** Habitat, behavior, and ecological traits were compiled from the literature (Nowak and Walker 1999) and the “Animal Diversity Web” (http://animaldiversity.ummz.umich.edu/site/index.html) (Myers et al. 2006). These data were used to define two ecological factors in terricolous mammals: spatial activity and rhythm activity. Rhythm activity was divided into three groups: diurnality, cathemeralism, and nocturnality. Nocturnal mammals had maximal activity during the dark period, whereas diurnal mammals had the reverse activity period (Kurumiya and Kawamura 1988). Cathemeral mammals included those that are approximately evenly active throughout the entire 24-h daily cycle (Tattersall 1979). For spatial activity, we classified mammals into two types, open-living behavior and nest-living behavior. Detail descriptions of ecological factors and citations for each species were summarized in supplementary table 1 (Supplementary Material online).

The Pearson product-moment correlation test, ANOVA, and ANOSIM were achieved using R programming language (http://www.r-project.org/). The Pearson product-moment correlation coefficient was based on covariance, and it gave information about the degree of correlation as well as the direction of the correlation as a linear relationship between two variables. Here, we compared the spatial activity and rhythm activity taken from empirical evidence with the number of intact V1Rs in 32 mammalian species that have different ecological factors. Rectangles represent different ecological factors including open-living behavior, nest-living behavior, diurnal activity, cathemeral activity, and nocturnal activity as labeled below each box. Error bars show the standard error of the mean. (A) Box plots showing the size of the intact V1Rs in 32 mammalian species. Median value and range of intact V1R gene numbers are shown. (B) Box plots showing the percentage of intact V1Rs in 32 mammalian species. Median value and range of the percentage of intact V1Rs are shown. Statistical significant differences are indicated by ** for $P < 0.01$ and *** for $P < 0.001$. $R$ represents the Pearson product-moment correlation coefficient.
intact V1R genes. We coded each type of rhythm activity as 1, 2, 3, corresponding to diurnality, cathemerality, and nocturnality, respectively. Open-living type and nest-living type behavior were coded 1 and 2 in spatial activity, respectively. ANOVA contained a collection of statistical models and their associated procedures, in which the observed variance was partitioned into components due to different explanatory variables. It gave a statistical test of whether the means of two or more groups were equal or not. The ANOSIM was a nonparametric test based on the rank ordering of the values of a distance matrix (e.g., Euclidean distance) among all observations and tested whether there was a statistically significant difference between two or more groups of sampling units or not. Here, we grouped mammals using different ecological factors. A total of 10,000 permutations were used to assess the significance of the ANOSIM test. The COMPARE 4.6b was applied to PICs analysis (Martins 2004). The topology and branch length of analyzed mammals were obtained from Ensembl genome database (http://www.ensembl.org/).

Supplementary Material
Supplementary material and table 1 are available at Genome Biology and Evolution online (http://www.oxfordjournals.org/our_journals/gbe/).

Acknowledgments
We thank David M. Irwin, Robert W. Murphy, and members of the Shi laboratory for valuable comments. This work was supported by grants from National Basic Research Program of China (973 Program, 2007CB411600), Chinese Academy of Sciences (KSCX2-YW-N-018), and Bureau of Science and Technology of Yunnan Province to Y-p. Z. and by a start-up fund of “Hundreds Talent Program” from Chinese Academy of Sciences and by grants from Key Project from National Natural Science Foundation of China (30930015) to P.S.

Literature Cited
Jaeger RG, Gergits WF. 1979. Intra- and interspecific communication in salamanders through chemical signals on the substrate. Anim Behav. 27:150–156.
Martins EP. 2004. COMPARE, version 4.6b. Computer programs for the statistical analysis of comparative data [Internet]. Department of Biology, Indiana University, Bloomington, Indiana. Available from: http://compare.bio.indiana.edu


**Associate editor:** George Zhang