Differences in Selection Drive Olfactory Receptor Genes in Different Directions in Dogs and Wolf

Rui Chen,1,2 David M. Irwin,1,3 and Ya-Ping Zhang1,4,*

1State Key Laboratory of Genetic Resources and Evolution, and Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China
2School of Life Science, University of Science and Technology of China, Hefei, China
3Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada
4Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming, China

*Corresponding author: E-mail: zhangyp@mail.kiz.ac.cn.
Associate Editor: John H. McDonald

Abstract

The olfactory receptor (OR) gene family is the largest gene family found in mammalian genomes. It is known to evolve through a birth-and-death process. Here, we characterized the sequences of 16 segregating OR pseudogenes in the samples of the wolf and the Chinese village dog (CVD) and compared them with the sequences from dogs of different breeds. Our results show that the segregating OR pseudogenes in breed dogs are under strong purifying selection, while evolving neutrally in the CVD, and show a more complicated pattern in the wolf. In the wolf, we found a trend to remove deleterious polymorphisms and accumulate nondeleterious polymorphisms. On the basis of protein structure of the ORs, we found that the distribution of different types of polymorphisms (synonymous, nonsynonymous, tolerated, and untolerated) varied greatly between the wolf and the breed dogs. In summary, our results suggest that different forms of selection have acted on the segregating OR pseudogenes in the CVD since domestication, breed dogs after breed formation, and ancestral wolf population, which has driven the evolution of these genes in different directions.

Key words: dog, wolf, polymorphism, purifying selection, tolerated, untolerated.

Introduction

Olfaction plays a large role in mammals, helping these animals discriminate noxious foods, mates, prey, marking territories, etc. (Firestein 2001; Mombaerts 2004). Olfaction is initiated by the binding of ligands to an olfactory receptor (OR) in the nasal cavity, with the OR gene family being the largest gene family in mammals (Glusman et al. 2001). Although mammals typically have a large number of OR genes, the number of intact genes varies considerably between species (Hayden et al. 2010). The previous studies have identified OR pseudogenes, and the generation of these genes have been correlated with evolutionary changes in specific groups. For example, in primates, the rapid and specific loss of OR genes (Gilad et al. 2003) is associated with the acquisition of trichromatic vision (Gilad et al. 2004). The fraction of the OR gene complements that are pseudogenes is determined by selective constraints (Rouquier et al. 2000), and these constraints have been found to vary between human populations (Gilad and Lancet 2003). Unlike Drosophila, where specific OR genes correspond to specific olfactory neurons (Ray et al. 2007), in mammals, the expression of OR genes is stochastically determined in a specific olfactory neuron, and an alternate OR gene will be activated if the first gene selected is a pseudogene (Shykind 2005). It is also known that specific amino acid residues in the OR sequences are required for guiding axons to the Glomeruli in the olfactory bulb (Feinstein and Mombaerts 2004).

The dog diverged from the wolf more than 10,000 years ago (Vila et al. 1997; Savolainen et al. 2002; Pang et al. 2009). Mitochondrial DNA studies indicate that the dog was domesticated from several hundred wolves in southern China (Pang et al. 2009). During dog evolution, two bottlenecks have been described; the first accomplished the domestication process, while the second occurred during the recent formation of the multiple modern breeds (Lindblad-Toh et al. 2005). The village dogs, including the Chinese village dog (CVD), are thought to represent an ancestral state of dog before breed formation and thus would not have experienced the second bottleneck associated with breed formation (Boyko et al. 2009; Boyko et al. 2010). Thus, analysis of village dogs should yield insight into the ancestral populations of dogs and, in comparison with modern breed dogs, the process of breed formation. After domestication, the dog has experienced both artificial and natural selections, where the selection on some genes has been relaxed (Bjørnerfeldt et al. 2006), whereas for others, it has been strengthened (Sutter et al. 2007). During breed formation, selection on some genes has been greatly strengthened by artificial selection (Vila et al. 1997; Vila et al. 2005; Mosher et al. 2007), and extensive linkage disequilibrium has been found in dog breeds (Sutter et al. 2004). As there have been two major bottlenecks in dog evolutionary history, most of the phenotypes of breed dogs can be explained by a small number of quantitative trait loci.
Different breeds.

The fisher's exact test was applied to determine whether the observed amino acid changes in the OR protein sequences are likely to be tolerated. Insertions and deletions were not considered. Due to the poor alignment of the N-terminal and C-terminal regions, we only considered sequences between amino acids 32 and 300, in accord with the previous study (Menashe et al. 2006). The J69 model was used to count the numbers of synonymous and nonsynonymous sites and tolerated and untolerated sites. (Jukes and Cantor 1969). If all the nonsynonymous polymorphisms between alleles at a specific position were tolerated, then we defined this polymorphism as a tolerated polymorphism. If at least one of the nonsynonymous polymorphisms at a position was found to be untolerated (damaging), then this is an untolerated polymorphism. The nucleotide diversity and numbers of synonymous and nonsynonymous sites were calculated with DnaSP 5.1 (Librado and Rozas 2009), and the raw data produced from SIFT were analyzed on an ABI PRISM 3730 DNA Analyzer (Applied Biosystems), and DNA sequences were edited using DNASTAR 5.0 (DNASTAR).

Data Acquisition

We focused on 16 segregating OR pseudogenes that were previously identified from a sample of 109 different OR genes in dogs of diverse breeds (the segregating OR pseudogenes are listed in table 8 of the study by Robin et al. [2009]) (supplementary table S1, Supplementary Material online). OR gene nomenclatures are from the study by Quignon et al. (2005). OR gene sequences from the wolves and CVD were obtained in our study, with haplotypes acquired through the use of PHASE (Scheet and Stephens 2006). The reference sequences were from the database of a previous study (http://dogs.genouest/ORreportoire.html) (Quignon et al. 2005); haplotypes and SNP frequencies for 16 segregating OR pseudogenes in dogs of different breeds were obtained from a previous study (Robin et al. 2009), and the sequences for breed dogs were recovered with a program written in C language.

Data Analysis

On the basis of the phylogenetic tree of OR genes (Quignon et al. 2005), we identified and obtained the sequences of the 29 most closely related genes of each of the segregating OR genes amplified in this study (supplementary table S2, Supplementary Material online). The 30 OR sequences, segregating pseudogene and 29 close relatives, were then aligned with MEGAS (Tamura et al., 2011). Alignments were then analyzed with SIFT (http://sift.jcvi.org) (Ng and Henikoff 2001) to predict whether the observed amino acid changes in the OR protein sequences are likely to be tolerated. Insertions and deletions were not considered. Due to the poor alignment of the N-terminal and C-terminal regions, we only considered sequences between amino acids 32 and 300, in accord with the previous study (Menashe et al. 2006). The J69 model was used to count the numbers of synonymous and nonsynonymous sites and tolerated and untolerated sites. (Jukes and Cantor 1969). If all the nonsynonymous polymorphisms between alleles at a specific position were tolerated, then we defined this polymorphism as a tolerated polymorphism. If at least one of the nonsynonymous polymorphisms at a position was found to be untolerated (damaging), then this is an untolerated polymorphism. The nucleotide diversity and numbers of synonymous and nonsynonymous sites were calculated with DnaSP 5.1 (Librado and Rozas 2009), and the raw data produced from SIFT were used to count the numbers of tolerated and untolerated sites. The Fisher's exact test was applied to determine whether the polymorphism ratios were significantly different from the site ratios. We divided the dataset into a wolf population, a CVD population, and the dog breed (dogs of different breeds) population. If a polymorphism existed in only one population,
then it was defined as a population-specific polymorphism, and thus, it was likely acquired after the foundation of that population. If a polymorphism exists in two populations, then it was defined as a polymorphism that was lost in the third population. Polymorphisms that were shared by more than one population were defined as ancestral polymorphisms.

The topology of the ORs proteins was predicted using HMMTOP (Tusnády and Simon 2001). The numbers of each type of polymorphism in the TM domains 1–7 (TM1–7), intracellular regions (ICRs), and ECRs were counted. The distributions of each type of polymorphism were defined as the number of that type in TM1–7, ICRs, and ECRs. As most polymorphisms in the NH2 and COOH tails are removed from the tolerated and untolerated polymorphisms analysis, we did not consider polymorphisms in the two tails during the polymorphism distribution analysis. We analyzed the distributions of polymorphisms using two approaches. First, polymorphisms were divided into four groups: ancestral polymorphisms, wolf-specific polymorphisms, CVD-specific polymorphisms, and breed dog-specific polymorphisms. In each group, the distributions of synonymous polymorphisms were compared with the distributions of the other types (nonsynonymous, tolerated, and untolerated polymorphisms). Second, the numbers of synonymous, nonsynonymous, tolerated, and untolerated polymorphisms and sites were counted in each of the three domains (TM1–7, ECRs, and ICRs) of the protein, and then the distributions of the numbers of polymorphisms of each type in the four groups (ancestral, wolf-specific, CVD-specific, and breed dog-specific polymorphisms) were compared with the distribution of the sites of that type. A chi-square test with 2 degrees of freedom was used to determine whether the distribution of synonymous, nonsynonymous, tolerated, and untolerated polymorphisms across groups is significantly different from distribution of sites or synonymous polymorphisms.

Tajima’s D (Tajima 1989), Fu and Li’s D*, Fu and Li’s F*, Fu and Li’s D, and Fu and Li’s F tests (Fu and Li 1993; Fu 1997) were used to test whether the genes in the different populations were evolving neutrally using DnaSP 5.1 (Librado and Rozas 2009).

**Results**

**Differences in the Levels of Gene Diversity between the Wolf and Dogs**

To examine gene diversity in dogs and wolves, we investigated 16 segregating OR pseudogenes that had been identified in a previous study (Robin et al. 2009). Gene sequences were obtained from all CVDs and wolves for each of the segregating OR pseudogenes except OR gene CfOR04c05 in two of the CVDs. Combining our new sequences with those from dog of different breeds, which were from a previous study (Robin et al. 2009), a total of 118 nonsynonymous and 85 synonymous polymorphisms and 14 indels (insertions and deletions) were found. Of the polymorphisms and indels identified, 65 polymorphisms and 1 indel were newly identified in the wolf and the CVD sequences (supplementary tables S3 and S4, Supplementary Material online). Of the 9 nonsense mutations and 13 indels identified in the previous study (Robin et al. 2009), 5 of the indels and 2 of the nonsense mutations were not found in either the wolves or the CVDs. The new indel that we found was a deletion in a CVD, and two of the new polymorphisms were nonsense mutations found in wolves (supplementary table S3, Supplementary Material online). Base C at position 702 of CfOR5912, from the OR gene repertoire (Quignon et al. 2005), is absent in all our new sequences. This base is present in the 1.5x Poodle genome (Kirkness et al. 2003) but absent in the 7.5x boxer genome (Lindblad-Toh et al. 2005) raising the possibility that this is a sequencing error and that this gene should be 900 bp long. Due to missing data after base 872, we have included only 872 bases of this gene in our analysis.

We also calculated the proportion of the alleles at the 16 segregating OR genes that were pseudogenes in the wolves, CVD, and 6 different breeds of dogs (table 1). The proportion of the alleles that were pseudogenes ranged from 0% to 100% for the different subpopulation, with average pseudogenes proportions ranging from 28.7% for the genes in CVD to 40.9% in the Labrador retriever. However, the differences in pseudogene proportions was not statistically significant different among the eight groups (Friedman test \( P = 0.629 \)).

The gene CfOR04c05 failed to produce an intact sequence from two individuals of CVD. In these two individuals, a PCR product of the expected size was generated for CfOR04c05, but the inner sequencing primer failed to generate a clear sequence due to the deletion of one base in one of the alleles in each individual. To calculate nucleotide diversity, we combined all gene sequences, except CfOR04c05. Wolves showed higher nucleotide diversity (0.00248) than CVD, who had intermediate nucleotide diversity (0.00228), whereas breed dogs showed much lower diversity than both other groups.

<table>
<thead>
<tr>
<th>OR Name</th>
<th>Wolf</th>
<th>CVD</th>
<th>BM</th>
<th>ESS</th>
<th>GREY</th>
<th>GSD</th>
<th>LR</th>
<th>PEK</th>
</tr>
</thead>
<tbody>
<tr>
<td>CfOR0004</td>
<td>0.58</td>
<td>0.42</td>
<td>0.63</td>
<td>0.56</td>
<td>0.88</td>
<td>0.88</td>
<td>0.56</td>
<td>0.56</td>
</tr>
<tr>
<td>CfOR0043</td>
<td>1.00</td>
<td>0.25</td>
<td>0.88</td>
<td>0.81</td>
<td>1.00</td>
<td>0.81</td>
<td>0.94</td>
<td>0.44</td>
</tr>
<tr>
<td>CfOR0135</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>CfOR0180</td>
<td>0.08</td>
<td>0.17</td>
<td>0.06</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>CfOR0401</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>CfOR0438</td>
<td>0.08</td>
<td>0.50</td>
<td>1.00</td>
<td>0.88</td>
<td>0.81</td>
<td>0.63</td>
<td>0.56</td>
<td>0.94</td>
</tr>
<tr>
<td>CfOR04C05</td>
<td>0.17</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.13</td>
</tr>
<tr>
<td>CfOR0519</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.94</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CfOR0565</td>
<td>0.08</td>
<td>0.00</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>CfOR0821</td>
<td>0.00</td>
<td>0.25</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.13</td>
</tr>
<tr>
<td>CfOR08G02</td>
<td>0.17</td>
<td>0.08</td>
<td>0.38</td>
<td>0.13</td>
<td>0.19</td>
<td>0.81</td>
<td>0.44</td>
<td>0.06</td>
</tr>
<tr>
<td>CfOR12F06</td>
<td>1.00</td>
<td>0.50</td>
<td>0.50</td>
<td>0.38</td>
<td>0.25</td>
<td>0.88</td>
<td>0.75</td>
<td>0.38</td>
</tr>
<tr>
<td>CfOR14A11</td>
<td>0.08</td>
<td>0.00</td>
<td>0.44</td>
<td>0.25</td>
<td>0.50</td>
<td>0.06</td>
<td>0.91</td>
<td>0.25</td>
</tr>
<tr>
<td>CfOR16C11</td>
<td>0.00</td>
<td>0.00</td>
<td>0.19</td>
<td>0.44</td>
<td>0.25</td>
<td>0.13</td>
<td>0.19</td>
<td>0.00</td>
</tr>
<tr>
<td>CfOR3109</td>
<td>0.75</td>
<td>0.92</td>
<td>1.00</td>
<td>0.63</td>
<td>0.56</td>
<td>1.00</td>
<td>0.88</td>
<td>1.00</td>
</tr>
<tr>
<td>CfOR5912</td>
<td>0.08</td>
<td>0.17</td>
<td>0.19</td>
<td>0.00</td>
<td>0.19</td>
<td>0.25</td>
<td>0.31</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Note.—Data for dog breed are from the haplotypes from a previous study (Robin et al. 2009). BM, Belgian Malinois; ESS, English Springer Spaniel; GREY, Greyhound; GSD, German Shepherd Dog; LR, Labrador Retriever; PEK, Pekingese.
showed the lowest diversity (0.00201) (table 2). These results are in accord with dogs representing a bottlenecked subpopulation of wolves, with breed dogs being further bottlenecked. These results are consistent with the CVD being more representative of an ancestral dog population before breed formation (Brown et al. 2011). When the ratio of nonsynonymous polymorphisms to synonymous polymorphisms was examined, it was found that wolves have the highest ratio, whereas breed dogs have the lowest ratio (supplementary table S4, Supplementary Material online).

Selection on the Different Types of Polymorphisms
To analyze whether the segregating OR genes are evolving neutrally in the different groups (wolves, CVDs, and breed dogs), we compared the ratios of nonsynonymous polymorphisms to synonymous polymorphisms with the ratio of nonsynonymous sites to synonymous sites. The ratio of nonsynonymous polymorphisms to synonymous polymorphisms is significantly lower than the ratio of nonsynonymous sites to synonymous sites for ancestral polymorphisms ($P < 0.01$, Fisher’s exact test one tailed) indicating that the OR genes in the ancestral populations of the three groups (wolves, CVD, and breed dogs) were not evolving neutrally and likely were functional (fig. 1). Population-specific polymorphisms of breed dogs also have a ratio of nonsynonymous polymorphisms to synonymous polymorphisms significantly lower than the ratio of nonsynonymous sites to synonymous sites ($P < 0.01$, Fisher’s exact test one tailed), thus purifying selection continued to occur in the breed dogs since their divergence from the village dogs.

Because some nonsynonymous mutations are not deleterious (tolerated), whereas some are deleterious (untolerated), the ratio of nonsynonymous polymorphisms to synonymous polymorphisms may not be powerful enough to detect positive selection, because at best only a small proportion of tolerated mutations are potentially under positive selection. Therefore, we divided the nonsynonymous polymorphisms into tolerated and un tolerated polymorphisms. As described in the Materials and Methods, we analyzed amino acid residues 32 to 300 of the OR sequences, where a total of 102 nonsynonymous polymorphisms were identified. Of the 102 polymorphisms, 40 generate un tolerated amino acid substitutions, whereas 62 generate tolerated amino acid substitutions (supplementary table S5, Supplementary Material online). A total of 4192 nonvariable amino acid sites (sites that do not have nonsynonymous polymorphisms) were identified from amino acid 32 to 300 (from amino acid 32 to 290 in CfORS912) of the intact copies of the 16 segregating pseudogenes. Of these 4192 nonvariable amino acid sites, 14 were identified as being un tolerated amino acids by SIFT (Ng and Henikoff 2001). If the intact copies of these segregating pseudogenes are indeed functional, then SIFT has incorrectly assessed the functional impact of these 14 residues. However, this result also indicates that the likelihood that functional residues being assessed as un tolerated by SIFT is less than 0.33% (14/4192). Further strengthening our conclusion that SIFT identified un tolerated mutations that impair function was the observation that the mutational direction of all the un tolerated substitutions was from a tolerated amino acid ancestor to an un tolerated amino acid derived sequence. Next, we compared the pairs of ratios. First, we compared the ratio of tolerated polymorphisms to synonymous polymorphisms with the ratio of tolerated sites to synonymous sites, and second, we compared the ratio of un tolerated polymorphisms to synonymous polymorphisms with the ratio of un tolerated sites to synonymous sites. For the ancestral polymorphisms, we found that ratios of tolerated polymorphisms to synonymous polymorphisms ($P < 0.01$, Fisher’s exact test one tailed) and un tolerated polymorphisms to synonymous polymorphisms ($P < 0.01$, Fisher’s exact test one tailed) are all significantly lower than the ratios of their sites (fig. 1). For polymorphisms that were specific to breed dogs, we also found that ratios of tolerated polymorphisms to synonymous polymorphisms ($P < 0.01$, Fisher’s exact test one tailed) and un tolerated polymorphisms to synonymous polymorphisms ($P < 0.01$, Fisher’s exact test one tailed) were significantly lower than the ratios of their sites (fig. 2). These observations suggest that selection is acting strongly against both tolerated and un tolerated polymorphisms within the different breeds of dog. In the specific polymorphisms of CVD and wolf, statistical tests were not significant for either the ratio of

---

**Table 2.** Test Statistics from the Neutrality Analysis of All Segregating OR Pseudogenes Except CfORS912.

<table>
<thead>
<tr>
<th></th>
<th>Tajima’s D</th>
<th>Nucleotide Diversity</th>
<th>Fu and Li’s D*</th>
<th>Fu and Li’s F*</th>
<th>Fu and Li’s D</th>
<th>Fu and Li’s F</th>
</tr>
</thead>
<tbody>
<tr>
<td>All samples</td>
<td>-0.40907</td>
<td>0.00232</td>
<td>-0.95052</td>
<td>-0.83822</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wolf</td>
<td>-0.43860</td>
<td>0.00248</td>
<td>-0.46486</td>
<td>-0.52246</td>
<td>-0.46892</td>
<td>-0.57499</td>
</tr>
<tr>
<td>CVD</td>
<td>-0.15074</td>
<td>0.00228</td>
<td>-0.20661</td>
<td>-0.21890</td>
<td>-0.15941</td>
<td>-0.19310</td>
</tr>
<tr>
<td>Breeds</td>
<td>0.72885</td>
<td>0.00201</td>
<td>0.58063</td>
<td>0.77230</td>
<td>0.72911</td>
<td>0.56698</td>
</tr>
<tr>
<td>BM</td>
<td>0.35535</td>
<td>0.00170</td>
<td>0.12812</td>
<td>0.22285</td>
<td>0.29702</td>
<td>0.23305</td>
</tr>
<tr>
<td>ESS</td>
<td>0.29364</td>
<td>0.00198</td>
<td>-0.06332</td>
<td>0.04436</td>
<td>-0.11361</td>
<td>0.03280</td>
</tr>
<tr>
<td>GREY</td>
<td>0.76230</td>
<td>0.00150</td>
<td>0.20054</td>
<td>0.41620</td>
<td>0.76024</td>
<td>0.97887</td>
</tr>
<tr>
<td>GSD</td>
<td>0.51345</td>
<td>0.00160</td>
<td>0.20012</td>
<td>0.33429</td>
<td>0.24500</td>
<td>0.36662</td>
</tr>
<tr>
<td>LR</td>
<td>0.45971</td>
<td>0.00181</td>
<td>0.53464</td>
<td>0.59341</td>
<td>0.49581</td>
<td>0.57840</td>
</tr>
<tr>
<td>PEK</td>
<td>0.46256</td>
<td>0.00170</td>
<td>-0.06292</td>
<td>0.10008</td>
<td>0.00942</td>
<td>0.16172</td>
</tr>
</tbody>
</table>

*NOTE.—Summary statistics for Tajima’s D, Fu and Li’s D*, Fu and Li’s F*, Fu and Li’s D, and Fu and Li’s F were calculated for the entire sample and the different populations (wolf, CVD, and different breeds of dog). BM, Belgian Malinois; ESS, English Springer Spaniel; GREY, Greyhound; GSD, German Shepherd Dog; LR, Labrador Retriever; PEK, Pekingese.
tolerated polymorphisms to synonymous polymorphisms or the ratio of un tolerated polymorphisms to synonymous polymorphisms, when compared with the ratios of their sites. However, for specific polymorphisms within wolves, the number of observed tolerated polymorphisms (19) was greater than expected (11.1), whereas the number of untolerated polymorphisms (8) was fewer than expected (14.6) under the expectation that the number of these polymorphisms should be proportional to the numbers of these sites. A similar observation is seen within CVD. These observations suggest that within CVD and wolves, there was a greater likelihood that tolerated changes were maintained, but that selection was still acting strongly against untolerated changes. The differences in the ratios of the different types of polymorphisms and their sites ratios in the three subpopulations (wolves, CVD, and breed dogs) suggest that different selection pressures are acting on the different subpopulations since their divergences.

The protein encoded by $\text{CFOR0180}$ was predicted by HMMP\textsc{top} to have only six rather than the expected seven TM domains; therefore, it was removed from our analysis of the distributions of polymorphisms. The distributions of nonsynonymous, tolerated, and untolerated polymorphisms in the ancestral polymorphisms were different from the distribution of synonymous polymorphisms, as fewer nonsynonymous, tolerated, and untolerated polymorphisms found in TM1–7 and more being located in the ECRs and ICRs of the protein. When the polymorphisms that were specific to the subpopulations were examined, differences were observed. In the wolf, the distribution of the nonsynonymous polymorphisms is different from that of the synonymous polymorphisms (fig. 3); however, in breed dogs, we found that the distributions of all four types of polymorphisms were almost the same. These observations suggest that selection against nonsynonymous, tolerated, and untolerated polymorphisms in TM1–7 has changed in breed dogs. Then the distributions of the ancestral, wolf-specific, CVD-specific, and breed dog-specific polymorphisms were compared with the sites distributions of the four polymorphic types (synonymous, nonsynonymous, tolerated, and untolerated polymorphisms) (fig. 4). Synonymous polymorphisms were found only in TM1–7 of CVD. For synonymous polymorphisms in the other three subpopulations (ancestral, wolf-specific, and breed-specific polymorphisms), no statistically significant difference was found. In the distribution of the other three types polymorphisms (nonsynonymous, tolerated, and untolerated polymorphisms), statistically significant ($P < 0.05$) differences were only found in the ancestral and wolf-specific polymorphisms, suggesting that selection has changed since domestication.

Neutral Test on Segregating Pseudogenes in Different Populations

To attempt to identify the type of selection that may have occurred in the three subpopulations, we applied Tajima’s D, Fu and Li’s D, and Fu and Li’s F tests. However, these tests can be influenced by population demographic dynamics in addition to selection. Negative values from these tests indicate an excess of rare alleles, which may be caused by positive selection, negative selection, or population expansion. Positive values indicate an excess of intermediate alleles, which may be due to balancing selection or a population bottleneck. Negative values for these tests were generated by the wolf and CVD data, whereas the breed dogs yield positive values (table 2). To distinguish selection from population dynamics, we separated nonsynonymous polymorphisms from synonymous polymorphisms, and for the nonsynonymous polymorphisms, we separated them into tolerated and untolerated polymorphisms (supplementary table S6, Supplementary Material online). Synonymous polymorphisms should be neutral, so that they can be used to detect population history. Negative values from these tests using synonymous polymorphisms for the wolf suggest that they are experiencing a population expansion, whereas positive value for the breed dogs suggests that they experienced a bottleneck. The test statistics for the CVD is near zero, which suggests that the CVD population size has remained nearly constant. Because population dynamics do not influence the test statistics in the CVD, in contrast to the wolf and breed dogs, we can then attribute the positive values from Tajima’s D, Fu and Li’s D, and Fu and Li’s F tests for balancing selection acting in the CVD population.
**Discussion**

**Nucleotide Diversities of Segregating OR Pseudogenes**

OR gene families adapt to the physical requirements of different species (Hayden et al. 2010), since olfaction is very important for mammals (Szetei et al. 2003). The ecological niche of dog has changed greatly since domestication from the wolf (Morey 1994); thus, it is likely that changes in the composition of the OR gene family have occurred between wolf and dog. The dog only recently diverged from the wolf, about 16,000 years ago (Pang et al. 2009), and there has been some backcrossing since domestication (Savolainen et al. 2002; Vila et al. 2005), which might explain why no significant difference in the OR pseudogene fraction between the wolf and dog was found (Zhang et al. 2011). In this study, we focused on 16 OR genes that had been previously shown to be segregating pseudogenes in breed dogs (Robin et al. 2009). We characterized these 16 genes in wolves, CVDs, and breed dogs and examined distribution patterns of polymorphisms to identify whether the selective forces acting on these sequences differ between the subpopulations.

The nucleotide diversities of the segregating OR pseudogenes, except CfOR04c05, were found to range from 0.00201 in the breed dogs to 0.00248 in wolves (table 2), values that are a little higher than the nucleotide diversity ($\theta = 0.0018$) calculated in the previous study of all 105 OR genes in breed dogs (Robin et al. 2009). Because the expected $\theta$ of genes based on the entire dog genome is 0.0005 (based on $\theta = 4N_eu$) (Lindblad-Toh et al. 2005), there may be three reasons for the elevated levels seen in the segregating OR pseudogenes. First, the necessity to interact with an enormous diversity of potential ligands may be the cause of the high nucleotide diversity of OR genes (Hayden et al. 2010). Second, balancing selection is known to be an important element in maintaining high nucleotide diversity, and this type of selection might be different between functional genes and segregating pseudogenes (Alonso et al. 2008). Third, the reduced effective population size of dog, especially breed dogs, may have led to lower nucleotide diversity in those dogs (Lindblad-Toh et al. 2005).

The gray wolf has experienced population contraction and expansion in the past 10,000 years (Lucchini et al. 2004), while
at the same time, the domesticated dog has gone through two bottlenecks due to domestication and breed formation (Lindblad-Toh et al. 2005). Negative values for synonymous polymorphisms are consistent with the expansion of the wolf population, whereas the near zero values indicate that the CVD population has remained nearly constant. Positive values for synonymous polymorphisms demonstrate that the breed dogs have experienced a bottleneck. Previous studies have shown that the bottleneck associated with breed formation was much more intensive than the bottleneck that occurred during domestication (Boyko et al. 2010); thus, the much lower values for nucleotide diversity observed within breeds in our data is in accord with this conclusion. However, population demography alone cannot explain the lower ratios of nonsynonymous polymorphisms to synonymous polymorphisms and suggests that selection continues to act on the segregating pseudogenes.

**Selection on the Segregating OR Pseudogenes**

OR genes in humans are maintained by balancing selection (Alonso et al. 2008). Higher values from the neutrality tests on nonsynonymous polymorphisms indicate that the segregating OR pseudogenes may be under balancing selection; however, the values was not significant, hence, we cannot make a strong conclusion. To refine our results, we divided the nonsynonymous polymorphisms into tolerated and untolerated polymorphisms based on their predicted effects on protein function. Untolerated substitutions are predicted to disrupt function (Ng and Henikoff 2001), whereas tolerated substitutions are not incompatible with function. However, accuracy of this analysis depends on the power of SIFT, and specific characters of OR genes may influence the results of this analysis. Because ORs must interact with a huge number of potential ligands in nature, tolerated polymorphisms in the extracellular domains may be beneficial for generating the diversity of OR function (Alonso et al. 2008). Discrimination of untolerated polymorphisms from tolerated polymorphisms in the nonsynonymous polymorphism analysis should allow a more accurate detection of selection, as purifying selection on the deleterious polymorphisms should reduce the ratio of nonsynonymous to synonymous polymorphisms. On the other hand, positive selection acting on beneficial polymorphisms would increase the nonsynonymous to synonymous ratio. With this analysis, by using the ratios of untolerated polymorphisms to synonymous polymorphisms and tolerated polymorphisms to synonymous polymorphisms, we can distinguish purifying from positive selection. In the wolf, selection is removing the untolerated polymorphisms and preserving the tolerated polymorphisms, whereas

---

**Fig. 3.** Distribution of polymorphisms in three domains of OR proteins (TM1–7, ICRs, and ECRs). Distribution of the four polymorphism types (synonymous polymorphisms [syn], nonsynonymous polymorphisms [nonsyn], tolerated polymorphisms [tole], untolerated polymorphisms [untole]) are shown for each domain. (A) Analysis of ancestral polymorphisms. (B) Analysis of specific polymorphisms of wolf. (C) Analysis of specific polymorphisms of CVD. (D) Analysis of specific polymorphisms of breed dogs. Significant differences in the patterns of distributions of sites to synonymous, nonsynonymous, tolerated, or untolerated polymorphisms were determined by a chi-square test with 2 degrees of freedom. P value of the statistical significance is shown above each significant type.
in the breed dogs, purifying selection is acting on both the unselected and selected polymorphisms. Gene conversion and recombination may help OR pseudogenes to become functional genes (Sharon et al. 1999), and our findings demonstrate that selection may also remove pseudogenes to preserve the functional genes.

Selection acts differently on the different domains of chemosensory receptors and on the orthologous genes in different species (Shi and Zhang 2006; Smadja et al. 2009). We analyzed the distribution patterns of polymorphisms across the different domains of the OR protein for the segregating OR pseudogenes and found there were significant differences between the distribution patterns of the synonymous polymorphisms with polymorphisms of other types (nonsynonymous, tolerated, and unselected) for the ancestral polymorphisms. However, in the subpopulation-specific polymorphisms, within breed dogs, the patterns of distribution for the three types of polymorphisms (nonsynonymous, tolerated, and unselected) are almost the same as the pattern of distribution for synonymous polymorphisms (fig. 3). In contrast, in the wolf, the patterns of distribution for the three types of polymorphisms differ from that for synonymous polymorphisms, with the wolf having a pattern more similar to the pattern of distributions for the ancestral polymorphisms (fig. 3). When the distribution patterns of the polymorphisms are compared with the patterns for the distribution of the sites, our statistical analysis did not find any significant differences for the CVD- or breed dog-specific (fig. 4), nonsynonymous, tolerated, and unselected polymorphisms. In contrast, the pattern of distributions for the wolf-specific and ancestral nonsynonymous and unselected polymorphisms are statistically significant ($P < 0.05$) or have a significant trend ($0.05 < P < 0.1$), that is, different from the distribution patterns for the sites, suggesting that the selection acting on these sites has changed since domestication.

In my previous analysis, a low ratio of nonsynonymous polymorphisms to synonymous polymorphisms in ancestral polymorphisms and specific polymorphisms of breed dogs was found. Our polymorphism distribution analysis suggests that the reasons for the similar patterns for ancestral and wolf polymorphisms compared with breed dogs are not identical. The low ratio for the ancestral polymorphisms may be explained by accelerated fixation of beneficial mutations due to positive selection and elimination of deleterious mutations by purifying selection. It is not simply the result of purifying selection. Accumulation of tolerated polymorphisms was also enhanced in wolves, with the number of unselected polymorphisms reduced, in accord with this conclusion. The low ratio observed in the specific polymorphisms of breed dogs is likely the result of artificial selection. The history
of most dog breeds is not more than 400 years (Parker et al. 2004). A great amount of linkage disequilibrium is found in the breed dogs (Sutter et al. 2004). Artificial selection, a form of purifying selection in our study, is not as complex as natural selections (Boyko et al. 2010) and may not be powerful enough to discriminate all tolerated and untolerated mutations. Our results suggest that selection on the OR genes has changed a great deal since the divergence of the dog breeds, whereas the analysis of the segregating OR pseudogenes in wolves indicates that they are still undergoing selection that favors functional genes.

**Conclusion**

Segregating pseudogenes exist during a subtle period of time, where they could become a functional or pseudogene within a population. In our investigation, we found that in natural populations, segregating OR pseudogene and pseudoalleles showed a trend to be eliminated in competition with functional alleles, whereas within domesticated dog, this trend has changed. In the CVD, these genes are evolving neutrally, whereas in breed dogs, they are evolving under artificial purifying selection. Genes can be beneficial in one period of time and then become deleterious in another period (Olson 1999; Zhang 2008). Our analysis demonstrates the selection acting on segregating OR pseudogenes has changed during the domestication of dogs, and thus, these genes are evolving in different ways among wolves, CVD, and breed dogs.

**Supplementary Material**

Supplementary tables S1–S6 and sequence information are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

The authors thank Guodong Wang for valuable advice and Stephanie Robin and Francis Galibert for their kind help. They also thank Dr John H. McDonald and anonymous reviewers for their comments and suggestions. This work was supported by grants from the National Basic Research Program of China (973 Program), Chinese Academy of Sciences, Bureau of Science and Technology of Yunnan Province, and National Natural Science Foundation of China.

**References**


Mosher DS, Quignon P, Bustamante CD, Sutter NB, Mellersh CS, Parker HG, Ostrander EA. 2007. A mutation in the myostatin gene


