Zoological Studies **48**(2): 223-237 (2009)

Zoological Studies

Evolution of Olfactory Receptor Genes in East Asian Loaches

Daryi Wang[1,*], Tzi-Yuan Wang[2], Li-Chuan Weng[1], Yasufumi Emori[3,4], Chyng-Shyan Tzeng[5], and Wen-Hsiung Li[1]

[1]*Biodiversity Research Center, Academia Sinica, Taipei 115, Taiwan*

[2]*Genomics Research Center, Academia Sinica, Taipei 115, Taiwan*

[3]*Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan*

[4]*Undergraduate Program for Bioinformatics and System Biology, University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan*

[5]*Institute of Bioinformatics and Structural Biology and Department of Life Sciences, National Tsing Hua University, Hsinchu 300, Taiwan*

(Accepted December 15, 2008)

Daryi Wang, Tzi-Yuan Wang, Li-Chuan Weng, Yasufumi Emori, Chyng-Shyan Tzeng, and Wen-Hsiung Li (2009) Evolution of olfactory receptor genes in East Asian loaches. *Zoological Studies* **48**(2): 223-237. It is well known that fish have fewer olfactory receptor (OR) genes than mammals do. In order to investigate the divergence of and evolutionary changes in OR genes in fish, 2 loach OR gene families (D1/3 and D32) from Taiwanese and Chinese populations of 2 species (*Misgurnus anguillicaudatus* and *Paramisgurnus dabryanus*) were analyzed. We found duplications in both D1/3 and D32 that had accumulated after the separation of the island of Taiwan from the Asian mainland ~5 million yrs ago. The accumulation of duplicate copies in a relatively short time suggests that gaining extra copies of OR genes may have been selectively advantageous for adapting to local environments. On the other hand, the loss of some OR genes in the 2 species in Taiwan may reflect loss of selective constraints on some OR genes. Moreover, unlike the situation in mammals in which pseudogenes are found in each family, pseudogenes are clustered in an expanding gene family in *P. dabryanus* in Taiwan. The presence of pseudogenes and the limited number of functional OR genes in fish suggest that fish do not need a large number of OR genes. Finally, the dynamic changes in OR gene number and the possible explanation for a lower OR gene number in fish are discussed.
http://zoolstud.sinica.edu.tw/Journals/48.2/223.pdf

Key words: Olfaction, Multigene family, Adaptation, *Misgurnus anguillicaudatus*, *Paramisgurnus dabryanus*.

Olfaction is a key sensory modality allowing individuals to detect environmental changes and to mediate behaviors such as mating, foraging, and escaping from danger. The ability of an organism to recognize and discriminate among odorants depends on the multiplicity and diversity of its olfactory receptors (ORs), encoded by a super gene family (Buck and Axel 1991). Currently, OR genes are placed into 2 classes (Glusman et al. 2000a). Class I contains the majority of fish OR genes, and it is thus also known as the fish-like OR class, whereas class II includes most mammalian OR genes. It was originally thought that class I genes were specialized to detect water-soluble odorants and class II genes could detect airborne odorants (Freitag et al. 1995). However, in recent studies numerous class I OR genes have also been found in humans and mice (Glusman et al. 2001, Zhang and Firestein 2002), and some class II OR genes have been found also in zebrafish (Niimura and Nei 2005b). A phylogenetic analysis of OR genes in fish, frogs, chickens, humans, and mice suggested that the common ancestor of fish and tetrapods possessed both types of OR genes

*To whom correspondence and reprints should be addressed. Tel: 886-2-27890159. Fax: 886-2-27827954. E-mail:dywang@gate.sinica.edu.tw

(Niimura and Nei 2005b).

Olfactory receptors are G protein-coupled receptors, which contain 7 α-helical transmembrane regions, which may interact with odorants. The size and diversity of OR families are used as an index for the discriminative ability of vertebrate olfactory systems. While as many as 1000 OR genes may exist in a mammal, in fish, the number ranges from only 40 to 150 (Buck and Axel 1991, Ben-Arie et al. 1994, Alioto and Ngai 2005). The difference in gene number is probably due to different adaptations to land and water. However, it remains a puzzle how fishes have survived with such a limited number of OR genes which detect water-soluble molecules, while some mammals may have over a 1000 OR genes to detect airborne odorants.

Evolutionary changes in OR gene numbers have been studied in humans, chimpanzees, and mice (Gilad et al. 2005, Niimura and Nei 2005a 2007). Humans possess many OR pseudogenes (Gilad et al. 2003). Niimura and Nei (2005a) estimated that the human lineage has lost ~430 functional OR genes, whereas the mouse lineage has gained ~350 new OR genes since the human-mouse split. The mouse-specific gene expansion was explained by their adaptation to diversified environments (Niimura and Nei 2005a). Although previous studies provided insights into the causes of evolutionary changes in OR gene numbers, most of them were restricted to model organisms with completely sequenced genomes (Glusman et al. 2001, Godfrey et al. 2004, Niimura and Nei 2005b, Nozawa and Nei 2007). In particular, there seems to be no study on closely related fish species or populations.

Most teleosts exclusively possess class I OR genes and have a much smaller OR gene family than do mammals. However, the sequence diversity in fish OR genes is higher than that in mammalian OR genes (Kratz et al. 2002). Therefore, fish OR genes are considered to be a good model to study the divergence and adaptation of the olfactory system. Two common loaches *Misgurnus anguillicaudatus* and *Paramisgurnus dabryanus*, also known as Oriental weatherfish, are widely distributed in South and East Asia, from Japan, Taiwan, and east China, to Myanmar. Both species dwell in rivers, lakes, and ponds and embed themselves in muddy bottoms with only their heads protruding for foraging. Therefore, it is widely believed that foraging loaches depend heavily on their chemosensory abilities. Both species are economic fish and can interbreed, thus implying that the 2 loaches are closely related, although morphologically distinct (Zhao et al. 2002, You et al. 2007). Recently, 24 OR genes of the Japanese loach (*M. anguillicaudatus*) were sequenced and classified into 4 subfamilies (Irie-Kushiyama et al. 2004). In this study, we selected 2 OR gene families (D1/3 and D32) in loaches to investigate how they evolved in response to life in different localities.

MATERIALS AND METHODS

Sample selection

Six *M. anguillicaudatus* and 8 *P. dabryanus* were collected from Taiwan and southern China in 2005. Five Japanese loach individuals were requested from Dr. Y. Emori (Irie-Kushiyama et al. 2004) (Fig. 1). In order to study the phylogenetic relationships among these loaches, the cytochrome b (*cytb*) gene sequences were obtained for the Taiwanese populations (see below) and the *cytb* sequences in Chinese and Japanese loaches were downloaded from GenBank, including those of *Cobitis striata*, *C. sinensis*, *M. fossilis*, and several Chinese populations of *M. anguillicaudatus* and *P. dabryanus*. Based on the phylogenetic analysis in figure 1, we selected 2 individuals (TW3 = TW, CH1 = CH) representing *M. anguillicaudatus* of Taiwan and Chinese populations, and 1 individual (TW4 = Pd) representing *P. dabryanus* for polymerase chain reaction (PCR) cloning of the OR genes.

DNA extraction, PCR amplification, and sequencing

For each sample, a 25 mg piece of pectoral or pelvic fin was immersed in 500 μl of digestion buffer (10 mM Tris-HCl (pH 8.0), 1% sodium dodecylsulfate, 2 mM ethylenediaminetetraacetic acid, 10 mM NaCl, 10 mg/ml dithiothreitol, and 0.5 mg/ml proteinase K) and incubated for 10-16 h at 50°C in a dry bath. Genomic DNA was isolated and purified with a DNeasy® Tissue Kit (QIAGEN, Hilden, Germany). The entire mitochondrial *cytb* gene was amplified and sequenced. The *cytb* gene segments (1141 base pairs) were amplified by a forward primer, PKG (5'-CCAGCGACTTGA AGAACCACCG-3'), and a reverse primer, G46 (5'-CTTTGGGAGCCAGGGGTGAG-3'), designed from the sequence of *Formosania lacustre*, a related species with a complete mitochondrial

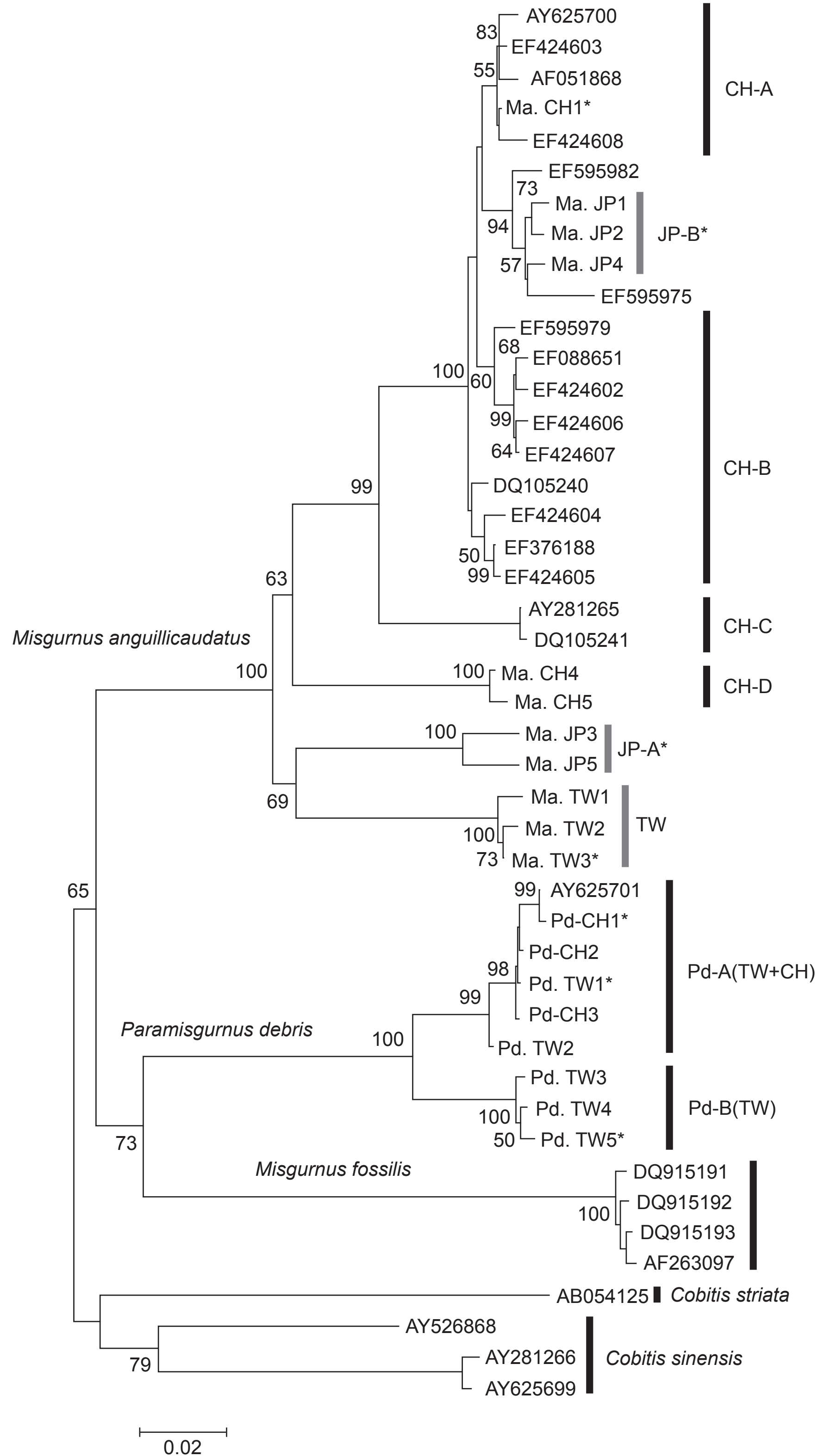

Fig. 1. Phylogenetic tree constructed using mitochondrion cytochrome b (*cytb*) gene sequences. *Cobitis* spp. (*C. striata* and *C. sinensis*) were used as outgroups. The tree was constructed using the Neighbor-joining (NJ) method with 1000 bootstrap replications. TW, Taiwanese population, CH, Chinese population, JP-A and JP-B: Japanese populations (Irie-Kushiyama et al. 2004). The samples used for the olfactory receptor (OR) gene cloning in this study are indicated by an asterisk (*).

genome. The PCR contained 50 ng of genomic DNA and supplementary material from the TaKaRa Ex Taq™ kit (Takara, Shinga, Japan) with standard PCR conditions. On average, about 1141 bp was sequenced for all samples in both directions using the sequence kits (BigDye™ Terminator Cycle Sequencing Ready Reaction Kits, Applied Biosystems, Foster City, CA, USA) and ABI model 377 automated DNA sequencers.

OR genes in the D32 and D1/3 families were amplified using degenerate primers and specific primer sets (Table 1). The PCR primers were designed according to the OR gene sequences obtained from Irie-Kushiyama et al. (2004), and they were selected with the Seqman software (DNASTAR, Madison, WI, USA). The degenerate primers are located in highly conserved regions, and the specific primers were exact matches to each subgroup of gene sequences. The PCR contained 1 μl (200 ng/μl) of genomic DNA, 5 μl of 10x Easy-A reaction buffer, and 2.5 U of the Easy-A High-Fidelity PCR cloning enzyme (Stratagene, La Jolla, CA, USA). The final dNTP concentration was 2.5 mM, with each primer at 10 μM. Each PCR included 28 cycles of denaturation at 95°C for 40 s, annealing at 50°C for the D-32 family and at 52°C for the D1/3 family for 1 min, and extension at 72°C for 1 min, with a final extension at 72°C for 7 min. Agarose-gel electrophoresis was performed to confirm the expected band size. The PCR product was then purified using a Qiagen purification kit and cloned with the yT&A vector cloning kit (Yeastern Biotech, Taipei, Taiwan). Plasmids were extracted using a Qiagen miniprep kit and sequenced as previously described.

Validation of OR genes

In order to thoroughly clone the genes in each of the 2 OR subfamilies being studied, over 200 clones were picked from each PCR with degenerate primers. The preparation and sequencing of each plasmid were done as described above. During the screening for OR gene sequences, only OR genes that appeared at least 5 times in the 200 selected clones were selected as true OR gene sequences, while others were considered PCR errors or recombinations. In addition, to determine whether single-copy clones were false positives or not, 3 clones (141, 198, and 214) obtained while cloning the TW-D32 family were randomly selected for verification. The primer sets used are listed in table 1.

PCR products utilizing stringent primers that were exact matches to sequences were cloned again, and over 30 clones were sequenced. Stringent PCR primers confirmed that some OR genes existed exclusively in 1 population.

Genomic real-time PCR

The relative OR gene ratios were analyzed using genomic real-time PCR to verify the accuracy of the OR gene numbers obtained from PCR cloning. Real-time PCR analyses were carried out in a final volume of 25 μl that contained 2 μl of diluted DNA (1 ng/μl), 1 μl of each gene-specific forward and reverse primers (5 μM), and 12.5 μl SYBR Green Supermix from Bio-Rad (Hercules, CA, USA) with 40 cycles of the following cycling program: denaturation at 95°C for 15 s and annealing at 64°C for 1 min. The primers were designed using the Primer Express software from Applied Biosystems. The relative copy number of OR gene was normalized using the D1/3-8 gene. The ΔCt relative quantification method was used in this study. The amplification efficiency of each primer pair was tested using 2-fold serial dilutions of the templates. ΔCt validation was carried out as suggested by the company, and the amplification efficiency of the target and reference genes were approximately equal. The D1/3-8 gene was selected as a standard because only 1 copy was obtained in each population; the CH-D1/3-1 subgroup was selected for the test. The primers used are listed in table 1.

Multiple sequence alignment and tree building

Sequences were aligned using ClustalX (Thompson et al. 1997). Modeltest vers. 3.7 (Posada and Crandall 1998) was used to select the substitution model that best described the data. The *cytb* gene was phylogenetically analyzed using the Neighbor-joining (NJ) method (Saitou and Nei 1987) with the Tamura-Nei model as implemented in MEGA 3.0 (Kumar et al. 2004). The reliability of the topology was evaluated by 1000 bootstrap replications.

The D1/3 and D32 family genes in the Japanese loach (*M. anguillicaudatus*) were downloaded from GenBank (GenBank accession nos. AB115055-AB115064, and AB115065-AB115072). The DNA sequences of the D1/3 and D32 families were phylogenetically analyzed separately using the NJ method with the Tamura-Nei model and 1000 bootstrap replicates.

The mean of sequence diversity was calculated using the modified Tamura-Nei method with 1000 bootstrap replicates in MEGA 3.0. The ratios of non-synonymous to synonymous substitution rates (*dN*/*dS*) of OR gene families in *M. anguillicaudatus* (TW3 = TW and CH1 = CH) and *P. dabryanus* (TW4 = Pd) were estimated using the method of Li (1993). In order to estimate the selective constraints on each OR gene, the *dN*/*dS* ratio for each branch was estimated using PAML 3.15 (Yang 1997). The ancestral nucleotide sequences were determined by first inferring the ancestral amino acid of each tree node and then inferring the ancestral nucleotide sequences under the constraint of the amino acid sequences inferred (Nei and Gojobori 1986, Zhang et al. 1998).

Gene conversion

The aligned DNA sequences were analyzed for gene conversion events using Sawyer's gene conversion detection program, GENECONV vers. 1.7 (Sawyer, 1999). The program infers whether pairs of sequences share unusual similarity. Global and pairwise *p* values are provided for sequence comparison in the program. Herein, we only used global *p* values from 10,000 permutations, because the author suggested that those values are more conservative and accurate. Default settings were used.

RESULTS

Phylogeny of *cytb* genes in loach populations

A phylogenetic tree of loaches was built using sequences of *cytb* genes. The tree topology shows that the sequence of *P. dabryanus* (Pd) was clearly separated from those of *M. anguillicaudatus* (Fig. 1); indeed, the genetic distance (Table 2) was as large as those among the 6 subspecies of Japanese loaches (Khan and Arai 2000). Therefore, it is reasonable to define *P. dabryanus* as a separate species living sympatrically with *M. anguillicaudatus*.

Using the average substitution number in the *cytb* gene between cluster TW and other CH clusters (108 substitutions) in *M. anguillicaudatus* and the average number between the 2 species (174 substitutions), we estimated a rough separation date of 8-9 million yrs (Myr) between the 2 species, assuming that the island of Taiwan separated from the Asian mainland ~5 Myr ago

Table 1. Degenerate primers for the D32 and D1/3 families and specific primers for each olfactory receptor (OR) subgroup

Primer name	Primer sequence	Primer name	Primer sequence
For OR gene cloning[a]		For single-copy clone test[b]	
D32-degenerate-F	ATGCCARBTTATTCMATGTGAA	TW-D32-9-F	TCATTATAGCCTTTGATAGG
D32-degenerate-R	TGGGATTTAACATTGGYGGRG	TW-D32-9-R	CTATAAGGTATAAACCTGTGTTA
D32-2-3-5-F	CAATTGTGCATCCTGAA	TW-D32-141-test-R	CTATAAGGTATAAACCTGTGCTG
D32-2-3-5-R	GATACGCTGATGATTCTT	TW-D32-5-F	GGACTTACGGGTTTACCATAT
D32-1-4-6-7-F	ATCCTGAATACTTTTTCATCGC	TW- D32-5-R	ATGGCATGATATCTGAGTGGC
D32-1-4-6-7-R	CATAAATAATGGGATTTAACAT	TW- D32-198-test-R	ATGGTGTGATATCTGAGTGGA
D32-8-F	TTTTGTCCGTCCTTGAATAC	TW- D32-2-F	TAATTCCTAACATGATGAGA
D32-8-R	AGAGATGCGCTGATAAT	TW- D32-2-R	AAACCTGTGTTAAGTTTTGCT
D1/3-degenerate-F	TT(AC)CGTCCDDMDACWTTYTTYA	TW- D32-214-test-R	AAACATGTGCTAAGTTGTCCC
D1/3-degenerate-R	WYWTTCACYYTSYTTCGTTTRTA	For genomic real time PCR	
D1/3-1-2-F	TAATGGCTTTTCTAATA	RT-D1/3-1F	AGTACCATTCCATCATAACCAAATCA
D1/3-1-2-R	GATTGCATGACCTCGTCTGTCT	RT- D1/3-1R	AACGAAAGTCTAGTAAGCAAACCTATGAC
D1/3-3-8-F	GTCTTTCACTTGGCTTTTTCTGA	RT- D1/3-8F	GCTATGAACACTTTAACGTCACATCTTT
D1/3-3-8-R	GATGGGATTCAACATTGGTGGTATT	RT- D1/3-8R	GGTTCAAGGATGTTGTAAAGCCA
D1/3-9-F	GTGCATTCAAACTCCTCTG		
D1/3-9-R	ATTGGTGGCATTGTCTGTGTCA		
D1/3-10-F	GTCCGTCCTGCTACATTCTTCATC		
D1/3-10-R	GGTGTTCATCCTGTTTCGTTTGTA		

[a] The number after the family name indicates the genes that exactly match the primer. [b] The 1st primer pair can amplify the OR gene, and the 3rd primer was designed to test the single-copy clones 141, 198, and 214.

(Huang et al. 2006, Shih and Suzuki 2008).

Cloning and validation of OR genes

To clone the OR genes, we selected 1 individual (TW4 = Pd) of *P. dabryanus* from the Taiwanese population, and 2 individuals (TW3 = TW and CH1 = CH) of *M. anguillicaudatus* respectively from a Taiwanese and Chinese population. We identified 10 genes and 1 pseudogene from TW, 10 genes from CH, and 11 genes and 1 pseudogene from Pd. For the loach D32 family genes, we identified 8 genes from TW, 9 genes from CH, and 9 genes and 2 pseudogenes from Pd. The numbers of cloned sequences for each OR family were similar to those obtained by genomic Southern blots of the D1/3 and D32 families of the Japanese loach (Irie-Kushiyama et al. 2004). Note that the sequence diversity among paralogous OR genes ranged 1.5%-7.7%, much higher than the average SNP frequencies of some teleosts (e.g., 0.69% in zebrafish; 0.43% in salmon; Woods et al. 2000, Smith et al. 2005). Therefore, the sequences cloned are considered to be family genes of different loci rather than genes of allelic heterogeneity. The new nucleotide sequence data were deposited in the DDBJ/EMBL/GenBank nucleotide sequence database (accession nos.: AB365356-AB365412).

Except for those OR genes that appeared at least 5 times in the 200 selected clones, all other sequences appeared only once in the 200 selected clones (data not shown). We assumed that these single sequences were produced by PCR errors or recombinations. Indeed, checking the sequence of each single-copy clone with the sequences derived from multiple clones, we found that 111 of 115 clones in the D1/3 family and all 136 clones in the D32 family could be explained by PCR recombination, and the remaining 4 single-copy clones in the D1/3 family could be explained as PCR mutations. To gain experimental support for this inference, we randomly selected 3 clones from the TW-D32 family, conducted a PCR with specific primers, and found that none of the sequences could be amplified (Fig. 2). Therefore, it is reasonable to infer that all of the single-copy clones were false positives.

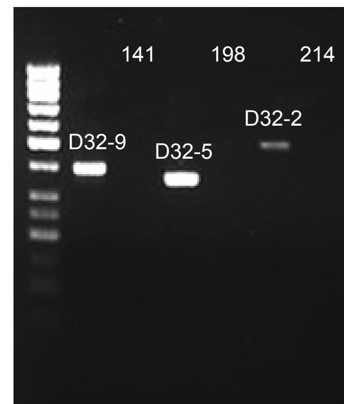

Fig. 2. Verification from polymerase chain reaction amplification. TW-D32-9, TW-D32-5, and TW-D32-2 were amplified using specific primers, while clones 141, 198, and 214 could not be amplified using specific primer sets.

Table 2. Net genetic distance and standard deviations between subgroups. For each between-group average, an arithmetic average was computed for all valid inter-group pairwise comparisons according to figure 1. The lower left matrix contains the net genetic distances; the upper right matrix has the standard deviations. *Misgurnus fossilis* (Mf) and two cobitids (Cs; *Cobitis striata* and *C. sinensis*) were used as outgroups in this study

	JP-A	JP-B	TW	CH	Pd	Mf	Cs
JP-A		0.0113	0.0101	0.0089	0.0141	0.0183	0.0097
JP-B	0.0998		0.011	0.0025	0.0154	0.0158	0.011
TW	0.0825	0.1086		0.0089	0.0149	0.0176	0.0101
CH	0.0747	0.0104	0.0854		0.014	0.0146	0.01
Pd	0.1641	0.1722	0.1712	0.1564		0.0151	0.011
Mf	0.2158	0.2019	0.2147	0.1847	0.1743		0.0137
Cs	0.1001	0.115	0.1092	0.1023	0.1119	0.1465	

JP-A, JP-B, two Japanese *Misgurnus anguillicaudatus*; TW, a Taiwanese *M. anguillicaudatus*; CH, a Chinese *M. anguillicaudatus*; Pd, *Paramisgurnus dabryanus*.

Further, in order to determine whether the OR gene numbers were reliable or not (e.g., an error due to allelic heterogeneity), a genomic real-time PCR was performed to estimate the copy number. We selected subgroup D1/3-1 for the test, and the results showed that the OR gene number from cloning matched the estimate from the genomic real-time PCR (Table 3).

Phylogenic analysis of OR families in loaches

Evolution of the D1/3 family

The phylogenetic tree of D1/3 family genes constructed by the NJ method (Fig. 3A) suggested 4 subgroups (α, β, γ, and δ), named according to Irie-Kushiyama et al. (2004). The fact that each Japanese (JP) loach sequence is clustered with TW, CH, or Pd sequences suggests that orthologous sequences of all known JP sequences were cloned in TW, CH, and Pd except for the JP-D1/3-4 and JP-D1/3-7 sequences. Note further that because gene conversion can occur between members of a gene family and because accelerated evolution can occur in some duplicate genes due to functional relaxation, the topology of the phylogenetic tree for a gene family might not be congruent with the species tree; that is, the tree for D1/3 family OR genes from these loaches might not be congruent with the tree in figure 1.

The α group in figure 3A contains 2 subgroups, although the bootstrap value for each group was very low, at only 42%. The tree topology for the 2nd subgroup was congruent with that in figure 1 for TW, CH, and Pd. TW D1/3-2a and TW D1/3-2b were clustered with CH-D1/3-2, and a simple explanation is that lineage-specific gene duplication occurred during the evolution of TW, and that TW-D1/3-2a underwent accelerated evolution following that duplication. The topology of the 1st subgroup was not congruent with that in figure 1 for TW, CH, and Pd. However, the 2 Pd (Pd-D1/3-1a and -1b) sequences are very similar to each other, so a simple explanation is that gene duplication occurred in Pd after its separation from TW-D1/3-1 and CH-D1/3-1. Therefore, there were probably 2 OR gene duplications in the α group. Although gene loss may also explain the observations, it is unlikely that many genes would be lost in a short time.

The β group contains 2 subgroups with low bootstrap values (39% and 32%). In the 1st subgroup, the topology for TW, CH, and Pd was congruent with that in figure 1, but the 2 Pd sequences are clustered, as are the 2 TW sequences and 2 CH sequences. This observation implies that 3 specific lineage duplication events occurred. However, we cannot exclude the possibility that the clustering of the 2 sequences in a taxa resulted from gene conversion. The 2nd subgroup contains 2 clades with a moderate (76%) and a low bootstrap value (28%), respectively. In both clades the tree topology for TW, CH, and Pd was congruent with that in figure 1, so there seemed to be no gene duplication.

The γ group contains 2 subgroups with high bootstrap values (95% and 98%). The 1st subgroup contains a single sequence each from Pd, TW, and CH, with a topology congruent with that in figure 1. Therefore, there was clearly no gene duplication during the evolution of this subgroup. The 2nd subgroup was divided into 2 clades, one with a single gene from TW, CH, and Pd, and the other with 1 gene from TW, 1 gene from CH, and 2 genes from Pd, suggesting gene duplication in Pd. These observations suggest that a gene duplication event occurred during the evolution of the γ group.

The δ group contains a single sequence from TW, CH, and Pd, so there is no evidence

Table 3. Ct values of the D1/3-1 and D1/3-8 genes detected by real-time polymerase chain reaction

	Ct of D1/3-1 Primer set	Ct of D1/3-8 Primer set	ΔCt	Olfactory receptor copy ratio(D1/3-1) / (D1/3-8)
TW	-	28.07 ± 0.48	-	1
CH	21.54 ± 0.37	21.67 ± 0.44	-0.13	1
Pd	25.86 ± 0.23	26.90 ± 0.11	-1.04	2

-, no data were detected. TW, a Taiwanese *Misgurnus anguillicaudatus*; CH, a Chinese *M. anguillicaudatus*; Pd, *Paramisgurnus dabryanus*.

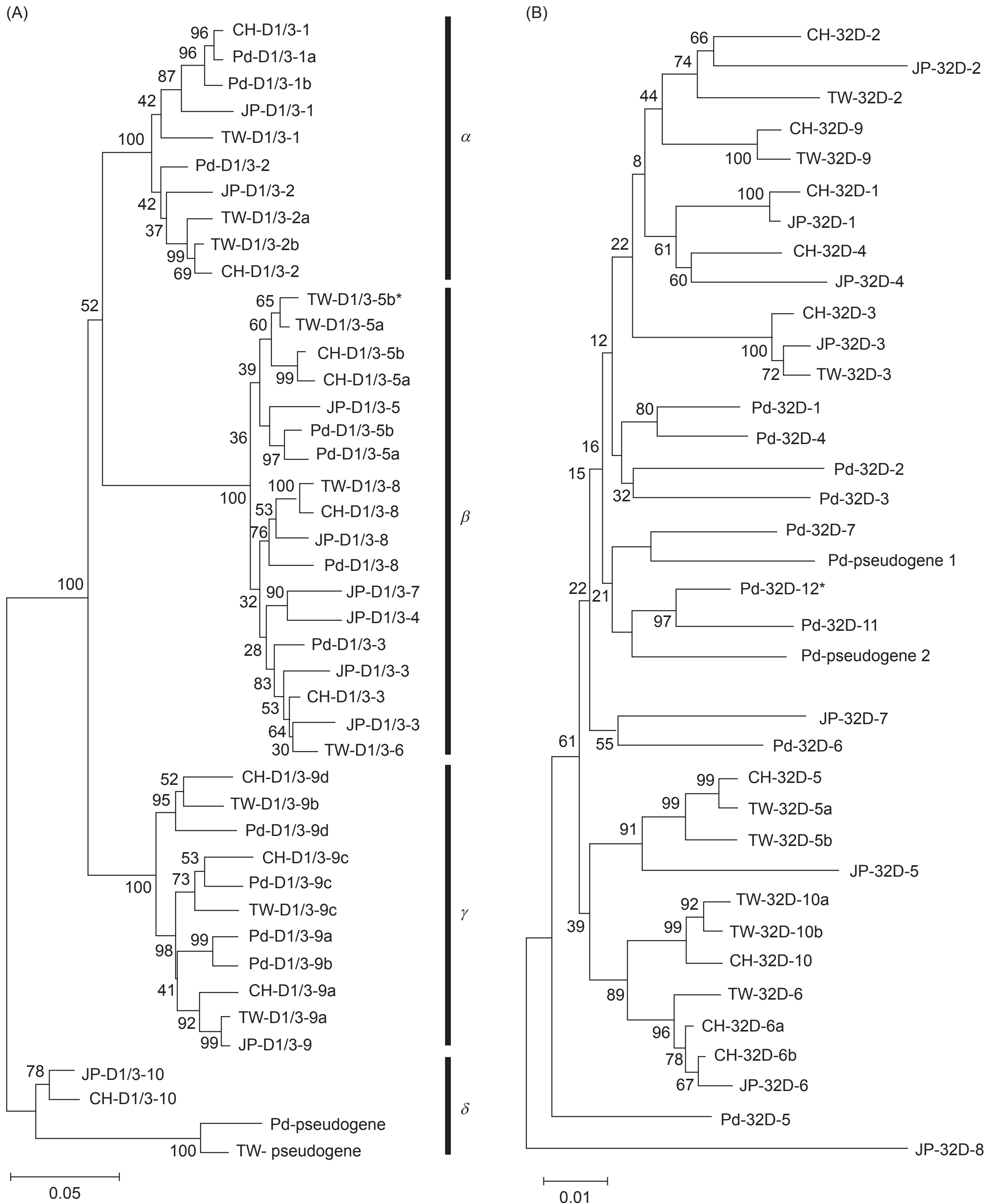

Fig. 3. Phylogenetic trees of D1/3 (A) and D32 (B) olfactory receptor genes in 2 loach species. The phylogenetic analysis was constructed using the Neighbor-joining method with the Tamura-Nei model. The number on a branch indicates the bootstrap support (proportion); 1000 bootstrap replications were conducted. Black bars indicate the 4 subtypes of the D1/3 gene. TW, *Misgurnus anguillicaudatus* from Taiwan; JP, *M. anguillicaudatus* from Japan (sequences quoted from Irie-Kushiyama et al. (2004)); CH, *M. anguillicaudatus* from China; and Pd, *Paramisgurnus dabryanus* from Taiwan. An asterisk (*) indicates a branch with *dN*/*dS* >1.

of gene duplication. The TW and Pd sequences are clustered together probably because, being pseudogenes, they have evolved faster than their functional counterparts in CH and JP.

In summary, the number of gene duplication events during the evolution of D1/3 OR genes was estimated to be between 3 and 9.

Evolution of the D32 family

In the phylogenetic tree of D32 genes (Fig. 3B), few of the major branching nodes have a high bootstrap value, so they should be used with caution. The upper part of the tree features a large cluster of Pd sequences (Pd-D32-1, -2, -3, -4, -6, -7, -11, and -12 and 2 pseudogenes). Taken at face value, this would suggest 5 or 6 duplication events during the evolution of Pd. However, the bootstrap values for most nodes are low, so one should allow for phylogenetic inference errors.

Phylogenetic inference errors may in part explain why few Pd sequences were clustered with sequences from other populations. However, 2 clusters, Pd-D32-1/-4 and Pd-D32-12/-11, had fairly high bootstrap values (80% and 97%), and we may infer there were at least 2 gene duplication events. In the lower part of the tree, there are 2 TW sequences (TW-D32-5a and -5b) in the 1st subcluster, 2 TW sequences (TW-D32-10a and -10b) in the 2nd subcluster, and 2 CH sequences (CH-D32-6a and -6b) in the 3rd subcluster. We can assume that each pair arose from gene duplication during the evolution of TW, CH, and Pd. In summary, there were probably 5 or more duplication events during the evolution of TW, CH, and Pd.

OR pseudogenes

In the D1/3 family, 1 TW pseudogene and 1 Pd pseudogene were clustered with JP D1/3-10 (the δ subgroup). The TW pseudogene included a 4 bp deletion that caused a frameshift and a premature stop codon. The Pd pseudogene included a 3 bp insertion and an 8 bp deletion, which caused a frame shift and a premature stop codon. The OR gene loss in the 2 subspecies (TW and Pd) might have occurred because it was no longer advantageous to maintain this gene in loach habitats of Taiwan.

In the D32 family, the JP-D32-8 gene is separate from all other sequences. We used D32-8-specific primers to amplify orthologous genes in TW, CH, and Pd, but failed to amplify any sequences. Checking the DNA alignment, we found that the JP-D32-8 gene possesses several signs of a pseudogene. One 2 bp deletion and two 1 bp insertions occurred between 17th and 36th nucleotides (a frame shifted in this fragment), along with some contiguous substitutions in the middle of the sequence. Thus, JP-D32-8 appears to be a pseudogene that has undergone rapid evolution, so that it is far separated from other D32 family members. Two other pseudogenes were found in Pd (Pd-pseudogenes 1 and 2) and were respectively clustered with Pd-D32-7 and Pd-D32-11. Pd-pseudogene 1 has a 24 bp deletion and 1 point mutation that caused a premature stop codon. Pd-pseudogene 2 has accumulated many mutations that resulted in frame shifts and premature stop codons. In view of the lineage OR gene expansions in Pd, the OR losses suggest that there was no advantage in expanding the OR gene number in this family.

Sequence diversity and substitution rate of OR genes

Multiple sequence alignments showed that the intracellular and extracellular regions of OR were relatively more conserved than the transmembrane (TM) domains in both the D1/3 and D32 families. Four leucines in TM3 and the inner domain connecting TM3 and TM4 and 4 cysteines in the extracellular loops between TM2 and TM3, TM4, and TM5 shared by all G protein-coupled receptors were highly conserved (Fig. 4A and 4B). In comparison, TM3, TM4, and TM5 were less well conserved.

The amino acid sequences of the D1/3 family showed an average high similarity of ~82.5% identity. However, levels of sequence diversity in some OR subgroups within these 3 loaches significantly differed ($p < 0.05$). Sequence diversities of the α and γ groups in the CH loach were 5.58 ± 0.80 and 8.28 ± 0.66, being 16.4%-39.1% higher than those in TW and Pd loaches (Table 4). This observation suggests that the CH loach harbors a larger pool of more-diversified D1/3 family members than do either the TW or Pd loaches. The higher diversity might be a result of adaptation to more-heterogeneous environments in China. A closer examination of the relative ratios of non-synonymous substitutions per site (*dN*) to synonymous substitutions per site (*dS*) in the D1/3 subfamily revealed that genes

(A)

```
#CH-D1/3-1   TFFINGFSNIPHGKYYYVFLSLVYAVTVFGNSFIMCIICLARRLHTAKYIVVFHLAFSDLCGSSALIPKLIDMFLFDNQYISYEACLANMFFVFHFMNLQSLTLLVLAFDRVIAICFPLKYHSIITKSTMFLIIAAMWIFSVAIFAIVIG
#CH-D1/3-2   ..................................................A..........................Y.................................L....L...........A.............T...VLV.
#TW-D1/3-1   .................A................................A..........................Y.................................L....L.....T.....A......V......SYVSV.V.
#TW-D1/3-2a  ..................................................A..........................Y.................................L....L.....T.....A......V......T....LV.
#TW-D1/3-2b  ..................................................A..........................Y.................................L....L.....T.....A.............T...VLV.
#Pd-D1/3-1a  .............................................................................Y.......................................L.....A.....A.....................
#Pd-D1/3-1b  ............................L.......................................................................................L.....A.....A.....V...............
#Pd-D1/3-2   ..................................................A..........................Y.................................M................A.............T....LV.
#JP-D1/3-1   ...........N...............E.................................................Y.................................L..........A.....A...................V.
#JP-D1/3-2   ..................................................A..........................Y..................I..............L..........A.....A.......I.....T....LV.
#CH-D1/3-5a  .......V.V..A...............L.......V.Y...........A..........................H.NVL..S..T...YIY...........A..Y..L.......R..A.V..TS.C.V..V......TY.SVYVS
#CH-D1/3-5b  ....................................V.Y...........A..........................H.NVL..S..T...YIY...........A..Y..L.......R..A.V..TS.C.V..V......TY.SVYVS
#CH-D1/3-8   ....................................V.Y...........A..........................H.NVL..S..T....IY....M......A..Y..L.......R..A.V..TS.C.V..V..M...TF.SLYVS
#TW-D1/3-3   ....................................V.Y...........A..........E...............H.NVL..S..T....IY....M......A..Y..L.......R..A.V..TS.C.V..V......TY.SVYVS
#TW-D1/3-5a  ....................................V.Y...........A..........................H.NVL..S..T...YIY...........A..Y..L.......R..DLV..TS.C.V..V......TY.SVYVS
#TW-D1/3-5b  .......V............................V.Y...........A..........................H.NVL..S..T...YIY...........A..Y..L.......R..A.V..TS.C.V..V......TY.SVYVS
#TW-D1/3-8   ....................................V.Y...........A..........................H.NVL..S..T....IY....M......A..Y..L.......R....V..TS.C.V..V......TY.SVYVS
#Pd-D1/3-3   ...............................L....V.Y...........A..........E...............H.NVL..S..T....IY....M......A..Y..L.......R..A.V..TS.C.V..V......TY.SGYVS
#Pd-D1/3-5a  .......V............................V.Y...........A..........................H.NVL..S..T....IY...........A..Y..L.......R..A.V..TS.C.V..V......TY.SGYVS
#Pd-D1/3-5b  .......V............................V.Y.......V...A..........................H.NVL..S..T....IY...........A..Y..L.......R..A.V..TS.C.V..V......TY.SVYVS
#Pd-D1/3-8   ....................................V.Y...........A........M.................H.NVL..S..T....IY...........A..Y..L.......R..A.V..TS.C.V..V......TF.SVYVS
#JP-D1/3-3   ....................................V.Y...........A..........E...............H.NVL..S..T....IY....M......A..Y..L.......R..A.V..TS.C.V..I..M...TY.SVYVS
#JP-D1/3-4   .......V............................V.Y...........A..........E...............H.NVL..S..T....IYL.L...........Y..L.......R..A.V..TS.C.V..V......TY.SVYVS
#JP-D1/3-5   .......V..............................Y...........A..........................H.NVL..S..T...YIY...........T..Y..L.......R..A.V..TS.C.V..V......TY.SVYVS
#JP-D1/3-6   ........................TL............Y...........A..........................H.NVL..S..T....IY....M......A..YG.L.......R..A.V..TS.C.V..V......TY.SVYVS
#JP-D1/3-7   .......V............................V.Y...........A..........................H.NVL..S..T....IY....M......A..Y..L.......R..A.V..TS.C.V..V......TY.SVNVS
#JP-D1/3-8   ....................................V.Y........R..A..........................H.NVL..S..T....IY....M......T..Y..L.......R..A.V..TS.C.V..V......RY.SVYVS
#CH-D1/3-9a  .......V.V..A.........................Y.......G..VA........M..........L.....EH....F.E..T.................S..Y..L...............R..C..MSV..T....F..VLV.
#CH-D1/3-9c  .......V.V.YA.......F.................Y..........VA........M..........L.....EH....F.E..T.................S..Y..L...............RA.C.L..V......TF..VLV.
#CH-D1/3-9d  .......V.V..A.........................Y...........A........M..........L.....EH....F.D....................S.SY..L...............R..C...F.......T...VLV.
#TW-D1/3-9a  .......V.V..A...............L.........Y..........VA........M..........L.....EH....F.D..T.................S..Y..L...............RA.C..M.V.....GTF..VLV.
#TW-D1/3-9c  .......V.V.YA.........................Y...........A........M..........L.....EH....F.D....................S..Y..L...............RA.C....V......TF..VLV.
#TW-D1/3-9d  .......V.V..A.......................V.Y.T.........A........M..........L.....EH....F.D..T.................S..Y..L...............R..C..M.V......T...VLV.
#Pd-D1/3-9a  .......V.V..A.................................G..VA........M..........L.....EH....F.D..T.L...............S..Y..L...............RF.C..M.V......TF..VLV.
#Pd-D1/3-9b  .......V.V..A.................................G..VA........M..........L.....EH....F.D..T.L...............S..Y..L...............RS.C..M.V......TF..VLV.
#Pd-D1/3-9c  .......V.V.YA.......F.................Y..........VA...................L.....EH...LF.D..T.................S..Y..L...............RS.C..M.V......TF..VLV.
#Pd-D1/3-9d  .......V.V..A...........V........................VA...................L.....EH....F.D....................S..Y..L...............QA.C...F.......T...VLV.
#JP-D1/3-9   .......V.V..A...............L.........Y..........VA........M..........L.....EH......D..T.................S..Y..L...............RA.C..M.V.....GTF..VLV.
#CH-D1/3-10  ....K....F.QA.......CV..TA..........V.Y..........VA......T.......V....L.....EH......NE.T.................A.....L....Y.....A.......SF..VG......L.L.VLA.
#JP-D1/3-10  ....K....F.QA.......CV.......W......V.Y..........VA......T.......V....L.....EH......NE.T.................A.....L....Y.....A.......SF..VG......F.L.VLA.
             ................................................50...............................................100...............................................150

#CH-D1/3-1   LLTRLSFCGSTVVNSYFCDHGPIYRLACNDNSLNQIMAHICFGLLMCLPLISIIISYVCIAVALLKITHGGDRTKAMKTCTSHLMLVAIFYLPSLSINILAITTSLNTNTRIINNSLSQTMPPMLNPIIYTLKTDEVMQSIKY
#CH-D1/3-2   ..................................N...N......L.....F.........................................G.....S.L...................S.....................
#TW-D1/3-1   .......................F..............N............F..................................................S........................................
#TW-D1/3-2a  ..........I.......................N...N............F.........................................G.....S.L...................S.....................
#TW-D1/3-2b  ..........I.......................N...N......L.....F.........................................G.....S.L...................S.....................
#Pd-D1/3-1a  ...................................................F...........................................................................................
#Pd-D1/3-1b  ..................................N................F.................................................L.........................................
#Pd-D1/3-2   ..................................N...N............F.........................................G.....S.L.........................................
#JP-D1/3-1   .......................F..............N............F...........................................................................................
#JP-D1/3-2   ..................................N...N............F.........................................G.....S.L.A.................S.....................
#CH-D1/3-5a  .VN.....RTN..D..Y.....VC......Y.I.VLFIKMT....V.....L.....S..SL..R..A...........L....L........Y......SF..TM.P.V.......T..I.............E.......D
#CH-D1/3-5b  .VN.....RTN..D..Y.....VC......Y.I.VLFIKMT....V.....L.....S..SL..R..A...........L....L........Y......SF..TM.P.V.......T..I.............E.......D
#CH-D1/3-8   .VN.....RTN..Y........V.........I.VLFIKMA....V.....L.....A..SL..H..A.........N.L....L........H..V...GF.....P.V.......T..I.............E.......D
#TW-D1/3-3   .VN.....RTNI.D..Y..I............I.VLFIKMS....L.....L........LL..R..A...........L....L........Y.....ISF..TM.P.V.......T..I.............E.......D
#TW-D1/3-5a  .VN.....RTN..D..Y.....VC......Y.I.VLFIKMT....V.....L.....S..SL..R..A...........L....L........Y......SF..TM.P.V.......T..I.............E.......D
#TW-D1/3-5b  .VN.....RTN..D..Y.....VC......Y.I.VLFIKMT....V.....L.....S..SL..R..A...........L....L........Y......SF..TM.P.V.......T..I.............E.......D
#TW-D1/3-8   .VN.....RTN..Y........V.........I.VLFIKMA....V.....L.....A..SL..H..A.........N.L....L........H..V...GF.....P.V.......T..I.............E.......D
#Pd-D1/3-3   FVN.....RTN..D..Y...............I.VLFIKMS....V.....L.....A..SL..R..A...........L....L........Y......SF..AM.P.V.......T..I.............E.......D
#Pd-D1/3-5a  .VN.....RTN..E..Y.....VCK.....Y.I.VLFIKMT....V.....L.....A..SL..R..A...........L....L........Y......SF..TM.P.V.......T..I.............E.......D
#Pd-D1/3-5b  .VN.....RTN..D..Y.....VC......Y.I.VLFIKMT....V.....L........SL..R..A...........L....L........Y......SF..TM.P.V.......T..I.............E.......D
#Pd-D1/3-8   .VN.....RTN...........V.I.......I.VLFIKMA....V.....L.....A..SL..R..A...........L....L........Y..H...GF.....P.V.......T..I.............E.......D
#JP-D1/3-3   FVN.....RTN..D..Y..I..........Y.I.VLFIKMS....L.....L........SL..R..A........I..L....L........Y.....ISF.KPM.P.V.......T..I.............E.......D
#JP-D1/3-4   WVN.....RTN..D..Y..I..........Y.IHVLFIKMT....V.....L........SL..R..G...........L....L...M....Y..H..VGF..PM.P.V.......T..I.............E.......D
#JP-D1/3-5   .IN.....RTN..D..Y.....VC......Y.I.VLFIKMT....V.....L........SL..R..A...........L....L........Y......SFA.TM.P.V.......T..I.............E........
#JP-D1/3-6   .VN.....KTN..D..Y..I............I.VLFIKMS....L.....L........SL..R..............L....L........Y......SF..TM.P.V.......T..I.............E.......D
#JP-D1/3-7   WVN.....KTN..D..Y..IS.........Y.ISVLFIKMT....V....VL.....S..SL..R..A...........L....L........Y..H..VGF.....P.V.V.....T..I.............E.......D
#JP-D1/3-8   .VN.....RTN..D..Y.....V.......Y.I.VLFIKMA....V.....L.....S..SL..R..A...........L....L........H..V...GF.....P.V.......T..I.............E.......D
#CH-D1/3-9a  ...............F....R.....S.......VM.GRM.....I....VL.....F...LV.F..V.NV.............L......I.I..N..A.AL..VT..VG..I...T........................D
#CH-D1/3-9c  ....................R...K.S.....K.VM.GRM.....I....VL.....F...LV.F..S.NV.............L......I.IF.N..A.AL.IHP..VEM.I...T...S............E.......V
#CH-D1/3-9d  ......................LH..G.....IHLM.GRM.....I....VL.....F...LV.F..S.NV.............L......I.L..N..A.AL.IHP..V.......A.I......................D
#TW-D1/3-9a  ...............F....R.....S.......VM.GRM.....I....VL.....F..GLV.F..V.NV.............L......I.I..N..A.AL..VT..VG..I...T...S............E.......D
#TW-D1/3-9c  ....................R...K.S.....K.V..GRM.....I....VL.....F..GLV.F..S.NV.............L......I.I..N..A.AL.IH...VE..I...A.I..............E.......D
#TW-D1/3-9d  ......................MH..G.....IHLM.GRM.....I....VL.....F..VLV.F..S.NV.............L......I.L..N..A.AL.IHP..V.......T........................D
#Pd-D1/3-9a  ....................R.....S.......IM.GRM.....I....VL.V...F...LV.F..A.NV.............L......I.V..N..A.ALA.VAP.VG..V...T...S............E.......D
#Pd-D1/3-9b  ....................R.....S.......IM.GRM.....I....AL.V...F...LV.F..A.NI.............L......I.V..N..A.AL..VVP.VG..I...T...S............E.......D
#Pd-D1/3-9c  .V..................R.....S.....K.LM.GRM.....I....VL.V...F...LV.F..S.NV.............L......I.I..N..A.ALA.VHP.VEM.I...T...S............E.......D
#Pd-D1/3-9d  ......................MH..G.....IKLM.GRM.....I....VL.....F...LV.F..S.NV.............L......I.L..N..A.AL.IHP..V.......A.I.................K....D
#JP-D1/3-9   ....................R.....S.......VM.GRM.....I....VL.....F..GLV.F..V.NV.............L......I.V..N..A.AL.--T..VG..I...T................E.......D
#CH-D1/3-10  .IN.....R..A..........VIS.S.....I.YLL.Y.S....I.M.VLLVT...L...A....LA..A..N..........A........I.V.VKSMS..IDPDA....T...T.II.............E.......D
#JP-D1/3-10  .IN.....R..A..........VIS.S.....I.YLL.Y.S....I.M.VLLVT...L...A....LA..A..N..........A........I..V.KSMS..IDLDS........T.II.............E.......D
             ...............................................200...............................................250...........................................
```

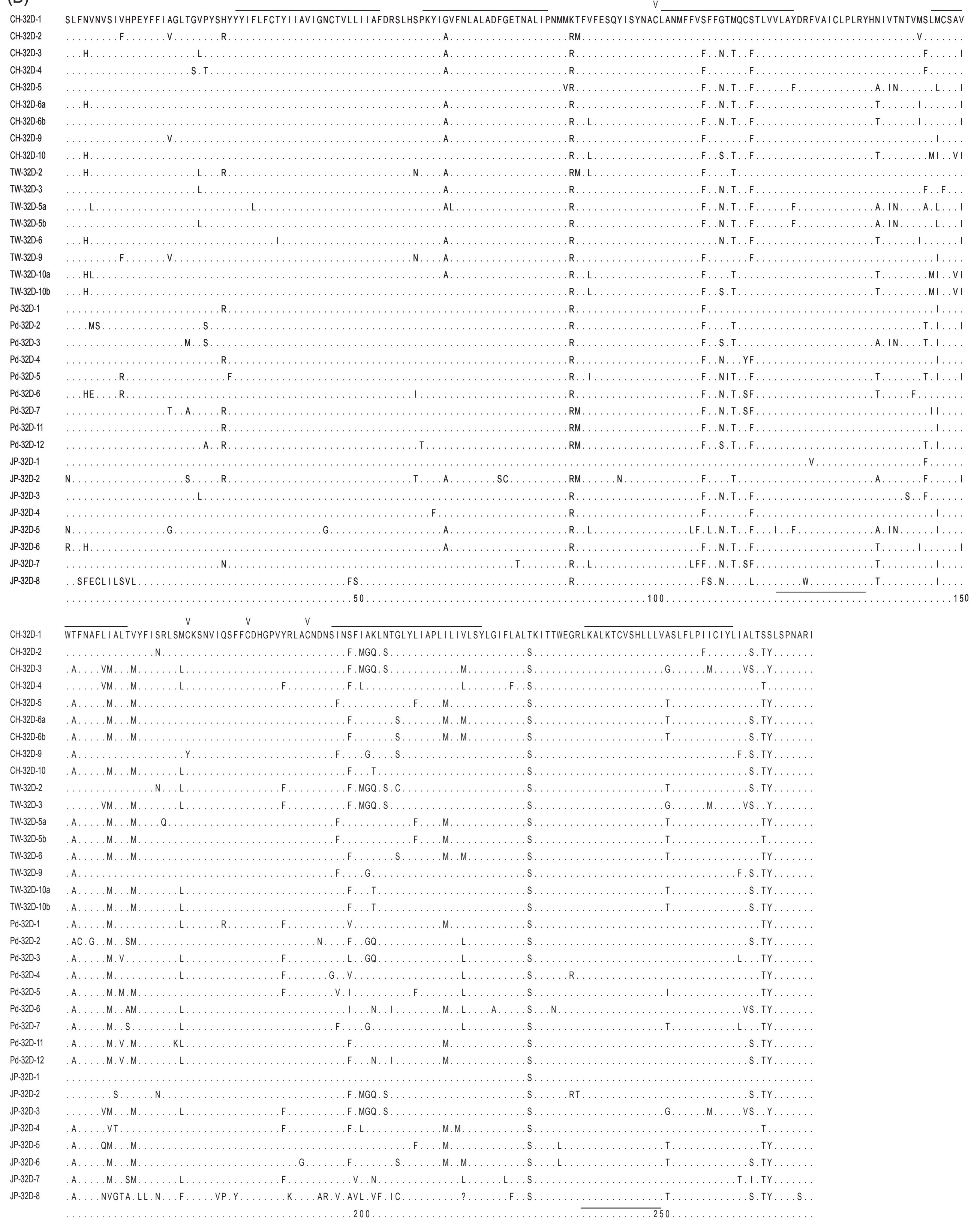

Fig. 4. Alignments of amino acid sequences encoded by genes in the D1/3 (A) and D32 families (B). The amino acid residues identical to the top sequence are marked with dots. Black lines indicate the predicted transmembrane (TM) domains. Gray lines indicate the conserved region shared by G protein-coupled receptors. Arrowheads indicate conserved Cys residues shared by olfactory receptor genes of vertebrates.

Table 4. Mean and standard deviation of sequence diversity (%) among olfactory receptor (OR) genes in each D1/3 subfamily or in the D32 family in a Taiwanese *Misgurnus anguillicaudatus* (TW), a Chinese *M. anguillicaudatus* (CH), and *Paramisgurnus dabryanus* (Pd)

OR genes	TW	CH	Pd
D1/3 α	4.10 ± 0.50	5.58 ± 0.80*	3.40 ± 0.49
D1/3 β	5.14 ± 0.61	4.25 ± 0.56	4.12 ± 0.05
D1/3 γ	6.92 ± 0.85	8.28 ± 0.66*	6.22 ± 0.68
D1/3 δ	-	-	-
D32	5.10 ± 0.49	4.72 ± 0.46	5.35 ± 0.49

*Indicates a significant difference between populations ($p < 0.05$) by *t*-test. -, Data insufficient for analysis.

Table 5. *dN*/*dS* of the olfactory receptor (OR) D1/3 subfamilies and D32 family in a Taiwanese *Misgurnus anguillicaudatus* (TW), a Chinese *M. anguillicaudatus* (CH), and *Paramisgurnus dabryanus* (Pd)

OR genes	TW	CH	Pd
D1/3 α	0.16 ± 0.07*	0.17 ± 0.07*	0.19 ± 0.03*
D1/3 β	0.30 ± 0.08	0.29 ± 0.04	0.26 ± 0.08
D1/3 γ	0.28 ± 0.06	0.26 ± 0.02	0.32 ± 0.12
D1/3 δ	-	-	-
D32	0.44 ± 0.11	0.45 ± 0.28	0.47 ± 0.7

*Indicates a significant difference ($p < 0.05$) in *dN*/*dS* between the OR subgroups by *t*-test. -, Data insufficient for analysis.

in the α group have been subjected to stronger purifying selection than those in the β or γ groups (Table 5). Only TW-D1/3-5b showed a *dN*/*dS* ratio of > 1 (1.15) in each branch (Fig. 3A).

The amino acid sequences of D32 family genes also showed a high similarity (90.6% on average). The level of sequence diversity (Table 4) and the *dN*/*dS* ratio (Table 5) showed no significant differences between each pair of subgroups (TW, CH, and Pd). However, *dN*/*dS* for the Pd-D32-12 branch had a high value of 2.2, significantly > 1 (Chi-test, $p < 0.001$) (Fig. 3B). In theory, a *dN*/*dS* ratio of > 1 indicates positive selection. The phenomenon of positive selection in the expanding cluster of Pd might reflect adaptations to the environment.

Putative gene conversion events

Six putative gene conversions were detected in the OR D1/3 family (Table 6). The predicted conversions were short (average 62 bp) compared to the total length of OR genes (~1000 bp). After removing these fragments, the tree topology remained the same as that in figure 3A, indicating no topological distortion by gene conversion. In the case of the D32 family, only 4 short converted fragments in the Pd population were predicted. After removing these fragments, Pd-D32-1, -2, -3, -4, -6, -7, -11, and -12 remained clustered together in the tree; that is, the 2 subgroups Pd-D32-1/-4 and Pd-D32-12/-11 were still supported by high bootstrap values (78% and 80%). Therefore, estimates of duplication events in the D32 family of Pd seemed to be largely reliable.

Table 6. Summary of detected conversion events of *Misgurnus anguillicaudatus* from Taiwan (TW) and China (CH) and *Paramisgurnus dabryanus* in Taiwan (Pd)

Convergent sequence pair		*p* value[a]	Seq. begin	Seq. end	Length
TW-D1/3-9c	TW-D1/3-1	0.0001	469	510	42
TW-D1/3-9a	TW-D1/3-1	0.0002	803	843	41
TW-D1/3-9a	TW-D1/3-3	0.0005	845	879	35
CH-D1/3-9d	CH-D1/3-1	0.0000	810	876	67
CH-D1/3-8	CH-D1/3-10	0.0001	820	851	32
CH-D1/3-1	CH-D1/3-5b	0.0003	1	108	108
Pd-D32-11	Pd-D32-2	0.0003	650	734	85
Pd-D32-4	Pd-D32-7	0.0004	81	256	176
Pd-D32-4	Pd-D32-7	0.0008	258	334	77
Pd-D32-4	Pd-D32-7	0.0003	391	432	42

[a]Fragments with global permutation *p* values of < 0.01 were selected.

DISCUSSION

Loach OR genes within a population showed only slight divergence. Four leucines in the TM3 and the inner domain connecting TM3 and TM4, and 4 cysteines in the extracellular loops between the TMs shared by all G protein-coupled receptors (Zhang and Firestein 2002, Liu et al. 2003) were conserved. In addition, the 2 landmark motifs (1 between TM3 and intracellular 2, and 1 between intracellular 3 and TM6) shared by most G protein-coupled receptors were also highly conserved (Zhang and Firestein 2002). Many features of OR sequences are shared with ORs of other vertebrates (Glusman et al. 2000b, Zhang and Firestein 2002, Liu et al. 2003). However, among the 21 predicted OR binding site positions (Man et al. 2004), 52% in the D1/3 family and 78% in the D32 family showed diversity. Interestingly, we observed a higher sequence diversity in the D1/3 α and γ subgroups of CH than in TW or Pd. As shown in figure 4, the 3 TM domains from China contained fewer conserved sequences and were predicted to play important roles in the potential ligand binding structure of OR genes (Pilpel et al. 1998, Liu et al. 2003). The higher sequence diversity in the CH population may reflect the potential for adaptation to larger river basins in China than in Taiwan.

Diversification of the TM of OR genes is commonly believed to be subject to positive selection for the ability to interact with diverse odorants (Gilad et al. 2000, Gilad et al. 2003). However, no positive selection was detected in the OR family genes of zebrafish (*Danio rerio*) or medaka (*Oryzias latipes*) (Sun et al. 1999, Alioto and Ngai 2005). It was suggested that human olfactory genes are under positive selection (Gilad et al. 2000, Gilad et al. 2003). However, some believe that the olfactory ability is largely determined by the number of genes, and that a small degree of selection would not be important (Niimura and Nei 2006, Niimura and Nei 2007). In our study, because we observed different selection pressures acting on different OR families, it was reasonable to infer that the OR families are responsible for sensing different essential odorants. In estimating *dN*/*dS* in each branch, we observed only 2 branches (out of 74 OR branches) with *dN*/*dS* values of > 1, and the remaining branches had *dN*/*dS* values of < 1. The observations of *dN*/*dS* values of > 1 may reflect adaptation to the environment. Although most OR genes are subject to purifying selection, the data we obtained on OR genes also suggested a process of rapid evolution (see below).

It was suggested that members of different subfamilies might recognize different classes or structural features of odorants, while members of the same subfamily might detect more-subtle differences between similar odorants (Buck and Axel 1991). Therefore, lineage-specific duplications may reflect an expansion in detection resolution. Current theories suggest 4 alternative outcomes for duplicated genes: (1) 1 copy may become nonfunctional; (2) 1 copy may acquire a new function and be preserved by natural selection; (3) the 2 genes may become specialized for different functions; and (4) the functions of both copies may become suboptimal (Ohno 1970, Hughes 1994, Force et al. 1999, Lynch and Conery 2000). The most common fate of duplicate genes is thought to be loss of a functional copy (Ohno 1970). According to the estimation using the sequence diversity of the *cytb* gene, *P. dabryanus* from Taiwan may have separated from *M. anguillicaudatus* 8-9 Myr ago. Taking D1/3 in 3 populations as an example, the average gene expansion was 0.023/gene/Myr (Demouth et al. 2006), higher than the average gene expansion in *Drosophila* (0.0012/gene/Myr) and in mammals (0.0016/gene/Myr) (Demuth et al. 2006, Hahn et al. 2007). Possible explanations for the homogenization of OR duplicate genes are gene conversion and duplication of tandem repeats (Kratz et al. 2002, Alioto and Ngai 2005). In either case, maintenance of duplicate copies might be beneficial. This is consistent with the observation that most duplicate copies had low *dN*/*dS* values, suggesting no sign of relaxation in selection pressure on duplicate genes.

OR gene families in a species reflect the ability of olfaction in that species. For example, OR genes in humans, mice, and canines possess different expansions in specific gene families (Olender et al. 2004, Gilad et al. 2005, Niimura and Nei 2007). Expansion within a family is commonly observed in mammals (Sosinsky et al. 2000). While as many as 1000 olfactory receptor genes might exist in a mammalian genome, the ORs in a fish species might be as few as 40-150. Our observations suggest that lineage-specific expansions have occurred. However, we also found that OR genes did not expand freely, because some of them became pseudogenes. TW and Pd loaches had pseudogenes in the δ group of D1/3, probably because they were no longer needed in the environments of Taiwan. In addition,

we found that gene gain (duplications) and loss (pseudogenization) concurrently occurred in the D32 family of the Pd population, which greatly differs from mammalian OR gene subfamilies that often expand largely within a cluster (Sosinsky et al. 2000, Niimura and Nei 2007). This observation suggests that Pd loaches do not need a wider spectrum in aquatic environments. The constraint in family size also explains why unlike mammals, fish have relatively small numbers of OR genes (Glusman et al. 2001, Godfrey et al. 2004, Gilad et al. 2005). In general, gene family size changes dynamically depending on the evolutionary lineage.

Acknowledgments: This work was supported by a research grant (96-2621-B-001-008-MY3) from the National Science Council of Taiwan to D. Wang.

REFERENCES

Alioto TS, J Ngai. 2005. The odorant receptor repertoire of teleost fish. BioMed Central Genomics **6:** 173.

Ben-Arie N, D Lancet, C Taylor, M Khen, N Walker, DH Ledbetter et al. 1994. Olfactory receptor gene cluster on human chromosome 17: possible duplication of an ancestral receptor repertoire. Hum. Mol. Genet. **3:** 229-235.

Buck L, R Axel. 1991. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell **65:** 175-187.

Demuth JP, T De Bie, JE Stajich, N Cristianini, MW Hahn. 2006. The evolution of mammalian gene families. PLoS ONE **1:** e85.

Force A, M Lynch, FB Pickett, A Amores, YL Yan, J Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151:** 1531-1545.

Freitag J, J Krieger, J Strotmann, H Breer. 1995. Two classes of olfactory receptors in *Xenopus laevis*. Neuron **15:** 1383-1392.

Gilad Y, O Man, G Glusman. 2005. A comparison of the human and chimpanzee olfactory receptor gene repertoires. Genome Res. **15:** 224-230.

Gilad Y, O Man, S Paabo, D Lancet. 2003. Human specific loss of olfactory receptor genes. Proc. Natl. Acad. Sci. USA **100:** 3324-3327.

Gilad Y, D Segre, K Skorecki, MW Nachman, D Lancet, D Sharon. 2000. Dichotomy of single-nucleotide polymorphism haplotypes in olfactory receptor genes and pseudogenes. Nat. Genet. **26:** 221-224.

Glusman G, A Bahar, D Sharon, Y Pilpel, J White, D Lancet. 2000a. The olfactory receptor gene superfamily: data mining, classification, and nomenclature. Mamm. Genome **11:** 1016-1023.

Glusman G, A Sosinsky, E Ben-Asher, N Avidan, D Sonkin, A Bahar et al. 2000b. Sequence, structure, and evolution of a complete human olfactory receptor gene cluster. Genomics **63:** 227-245.

Glusman G, I Yanai, I Rubin, D Lancet. 2001. The complete human olfactory subgenome. Genome Res. **11:** 685-702.

Godfrey PA, B Malnic, LB Buck. 2004. The mouse olfactory receptor gene family. Proc. Natl. Acad. Sci. USA **101:** 2156-2161.

Hahn MW, MV Han, SG Han. 2007. Gene family evolution across 12 *Drosophila* genomes. PLoS Genet **3:** e197.

Huang CY, PB Yuan, SJ Tsao. 2006. Temporal and spatial records of active arc-continent collision in Taiwan. Bull. Geol. Soc. Am. **118:** 274-288.

Hughes AL. 1994. The evolution of functionally novel proteins after gene duplication. Proc Biol Sci **256:** 119-124.

Irie-Kushiyama S, M Asano-Miyoshi, T Suda, K Abe, Y Emori. 2004. Identification of 24 genes and two pseudogenes coding for olfactory receptors in Japanese loach, classified into four subfamilies: a putative evolutionary process for fish olfactory receptor genes by comprehensive phylogenetic analysis. Gene **325:** 123-135.

Khan MR, K Arai. 2000. Allozyme variation and genetic differentiation in the loach *Misgurnus anguillicaudatus*. Fish. Sci. **66:** 211-222.

Kratz E, JC Dugas, J Ngai. 2002. Odorant receptor gene regulation: implications from genomic organization. Trends Genet. **18:** 29-34.

Kumar S, K Tamura, M Nei. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform. **5:** 150-163.

Liu AH, X Zhang, GA Stolovitzky, A Califano, SJ Firestein. 2003. Motif-based construction of a functional map for mammalian olfactory receptors. Genomics **81:** 443-456.

Lynch M, JS Conery. 2000. The evolutionary fate and consequences of duplicate genes. Science **290:** 1151-1155.

Man O, Y Gilad, D Lancet. 2004. Prediction of the odorant binding site of olfactory receptor proteins by human-mouse comparisons. Protein Sci. **13:** 240-254.

Nei M, T Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:** 418-426.

Niimura Y, M Nei. 2005a. Evolutionary changes of the number of olfactory receptor genes in the human and mouse lineages. Gene **346:** 23-28.

Niimura Y, M Nei. 2005b. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. Proc. Natl. Acad. Sci. USA **102:** 6039-6044.

Niimura Y, M Nei. 2006. Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates. J. Hum. Genet. **51:** 505-517.

Niimura Y, M Nei. 2007. Extensive gains and losses of olfactory receptor genes in mammalian evolution. PLoS ONE **2:** e708.

Nozawa M, M Nei. 2007. Evolutionary dynamics of olfactory receptor genes in *Drosophila* species. Proc. Natl. Acad. Sci. USA **104:** 7122-7127.

Ohno S, ed. 1970. Evolution by gene duplication. Heidelberg: Springer-Verlag.

Olender T, T Fuchs, C Linhart, R Shamir, M Adams, F Kalush et al. 2004. The canine olfactory subgenome. Genomics **83:** 361-372.

Pilpel Y, A Sosinsky, D Lancet. 1998. Molecular biology of olfactory receptors. Essays Biochem. **33:** 93-104.

Posada D, KA Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14:** 817-818.

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

Sawyer SA. 1999. GENECONV: A computer package for the statistical detection of gene conversion. Distributed by the author, Department of Mathematics, Washington University, St. Louis, MO, USA. Available at http://www.math.wustl.edu/~sawyer

Shih HT, H Suzuki. 2008. Taxonomy, phylogeny, and biogeography of the endemic mudflat crab *Helice/Chasmagnathus* complex (Crustacea: Brachyura: Varunidae) from East Asia. Zool. Stud. **47:** 114-125.

Smith CT, CM Elfstrom, LW Seeb, JE Seeb. 2005. Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. Mol. Ecol. **14:** 4193-4203.

Sosinsky A, G Glusman, D Lancet. 2000. The genomic structure of human olfactory receptor genes. Genomics **70:** 49-61.

Sun H, R Kondo, A Shima, K Naruse, H Hori, SI Chigusa. 1999. Evolutionary analysis of putative olfactory receptor genes of medaka fish, *Oryzias latipes*. Gene **231:** 137-145.

Thompson JD, TJ Gibson, F Plewniak, F Jeanmougin, DG Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:** 4876-4882.

Woods IG, PD Kelly, F Chu, P Ngo-Hazelett, YL Yan, H Huang et al. 2000. A comparative map of the zebrafish genome. Genome Res. **10:** 1903-1914.

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

You C, X Yu, J Tong. 2007. Detection of hybridization between two loach species (*Paramisgurnus dabryanus* and *Misgurnus anguillicaudatus*) in wild populations. Environ. Biol. Fish. (in press)

Zhang J, HF Rosenberg, M Nei. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA **95:** 3708-3713.

Zhang X, S Firestein. 2002. The olfactory receptor gene superfamily of the mouse. Nat. Neurosci. **5:** 124-133.

Zhao ZS, GQ Gao, QJ Wu. 2002. The chromosome set variation in hybridization of *Misgurnus anguillicaudatus* with *Paramisgurnus dabryanus* and in artificial and rogenetic *P. dabryanus*. J Dalian Fish. Univ. **17:** 15-19. (in Chinese)