# Genomic Organization and Evolution of the Vomeronasal Type 2 Receptor-Like (*OlfC*) Gene Clusters in Atlantic Salmon, *Salmo salar*

*Kimberley A. Johnstone,* Kate L. Ciborowski,† Krzysztof P. Lubieniecki,* William Chow,* Ruth B. Phillips,‡ Ben F. Koop,§ William C. Jordan,† and William S. Davidson**

*Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada; †Institute of Zoology, Zoological Society of London, Regent's Park, London, UK; ‡Department of Biological Sciences, Washington State University-Vancouver; and §Department of Biology, University of Victoria, Victoria, BC, Canada

There are three major multigene superfamilies of olfactory receptors (OR, V1R, and V2R) in mammals. The ORs are expressed in the main olfactory organ, whereas the V1Rs and V2Rs are located in the vomeronasal organ. Fish only possess one olfactory organ in each nasal cavity, the olfactory rosette; therefore, it has been proposed that their V2R-like genes be classified as olfactory C family G protein-coupled receptors (*OlfC*). There are large variations in the sizes of OR gene repertoires. Previous studies have shown that fish have between 12 and 46 functional V2R-like genes, whereas humans have lost all functional V2Rs, and frog sp. have more than 240. Pseudogenization of V2R genes is a prevalent event across species. In the mouse and frog genomes, there are approximately double the number of pseudogenes compared with functional genes. An oligonucleotide probe was designed from a conserved sequence from four Atlantic salmon O*lfC* genes and used to screen the Atlantic salmon bacterial artificial chromosome (BAC) library. Hybridization-positive BACs were matched to fingerprint contigs, and representative BACs were shotgun cloned and sequenced. We identified 55 O*lfC* genes. Twenty-nine of the *OlfC* genes are classified as putatively functional genes and 26 as pseudogenes. The *OlfC* genes are found in two genomic clusters on chromosomes 9 and 20. Phylogenetic analysis revealed that the *OlfC* genes could be divided into 10 subfamilies, with nine of these subfamilies corresponding to subfamilies found in other teleosts and one being salmon specific. There is also a large expansion in the number of *OlfC* genes in one subfamily in Atlantic salmon. Subfamily gene expansions have been identified in other teleosts, and these differences in gene number reflect species-specific evolutionary requirements for olfaction. Total RNA was isolated from the olfactory epithelium and other tissues from a presmolt to examine the expression of the odorant genes. Several of the putative *OlfC* genes that we identified are expressed only in the olfactory epithelium, consistent with these genes encoding odorant receptors.

## Introduction

The homeward migration of salmonids is reliant on chemical cues imprinted during their migration from freshwater to the marine environment. These chemical landmarks aid salmon to navigate back to their natal streams, which is an important ecological and evolutionary phenomenon because it leads to genetically distinct populations (King et al. 2001). To understand the molecular basis for homing in salmon, it is important to know how many olfactory receptor genes there are in salmonid genomes and how they are organized and expressed.

Olfactory receptor (OR) genes belong to what is considered to be the largest multigene superfamily within mammalian genomes (Alioto and Ngai 2005). These ORs are G protein-coupled receptors (GPCRs) that are characterized by seven α-helical transmembrane domains (Niimura and Nei 2005). Mammals have two olfactory organs, the main olfactory organ and the vomeronasal organ, whereas fish have a single olfactory organ in each nasal cavity, the olfactory rosette. In mammals, the main ORs are expressed in the ciliated neurons of the main olfactory epithelium, whereas the vomeronasal receptors (VNRs) are expressed in the microvillar cells of the vomeronasal organ. The human (*Homo sapiens*) genome contains ~800 OR genes, ~50% of which are pseudogenes, whereas the mouse (*Mus musculus*) genome has ~1,400 OR genes, with ~25% of them being pseudogenes (Niimura and Nei 2005). In contrast, it has been estimated that teleost genomes contain far fewer OR genes, with 143 in zebra fish (*Danio rerio*) and 42–44 in green spotted pufferfish (*Tetraodon nigroviridis*) (Alioto and Ngai 2005).

The mammalian VNR family is subdivided into two types: the vomeronasal receptor family 1 (V1R) and vomeronasal receptor family 2 (V2R). Fish do not have a vomeronasal system, and their corresponding ORs are expressed in the olfactory epithelium of the nasal cavity. The fish V1R-like and V2R-like receptors were originally named in accordance with the mammalian nomenclature, but it has recently been proposed to name the V1R-like genes as *ora* (ORs related to class A GPCRs) (Saraiva and Korsching 2007; Johnstone et al. 2008) and the V2R-like genes as *Olf*C (ORs related to class C GPCRs) (Alioto and Ngai 2006). The members of the OlfC family have long N-terminal extracellular domains that are used for initiating ligand binding (Han and Hampson 1999). Other members of this family include metabolic glutamate receptors (mGluR), extracellular calcium sensing receptors (CaSR), and gamma-aminobutyric acid receptors. There are large variations in the sizes of V2R and OlfC repertoires. Humans appear to have lost all functional V2Rs, whereas frog sp. have 249 V2Rs, and fish have between 12 and 46 *Olf*C genes (Hashiguchi and Nishida 2006; Shi and Zhang 2007). Pseudogenization of V2R genes is a prevalent event across species. In the mouse and frog (*Xenopus tropicalis*) genomes, there are approximately double the number of pseudogenes compared with functional genes (Shi and Zhang 2007), whereas in fish, the percent of pseudogenziation ranges from 19% to 47% of all *Olf*C genes (Hashiguchi and Nishida 2006). *OlfC* genes are found in large genomic clusters that vary in size depending upon the teleost species that have been examined. Medaka (*Oryzias latipes*) and green spotted pufferfish each have a single *OlfC* cluster that covers less than 300 kb. In contrast, zebra fish has two genomic clusters covering 4 Mb of the

Key words: olfactory receptor, G protein-coupled receptor, salmonids.

E-mail address: wdavidso@sfu.ca.

*Mol. Biol. Evol.* 26(5):1117–1125. 2009
doi:10.1093/molbev/msp027
Advance Access publication February 12, 2009

**Table 1**
**Hybridization and PCR Primers for the Identification of *OlfC* Genes in the Atlantic Salmon BAC Library**

| Primer Name | Sequence (5′–3′) |
|---|---|
| SVRA R probe | CAGGAGTCCAATATAGCCCA ACACAGCCCAGAAGCCAATA |
| DQ375533 SVRA F | CTTTCACTCTCATACAGGTC |
| DQ375535 SVRA F | CCTTCACGTTTGTCCAGGCT |
| DQ375537 SVRB F | GCTGTACCCTCCTACAGGTG |

genome, and fugu (*Takifugu rubripes*) appears to have four clusters, which may be because the genome has not been fully assembled.

The function of *OlfC* genes in Atlantic salmon (*Salmo salar*) is unknown. However, two orthologues in goldfish (*Carassius auratus*) and zebra fish (receptor 5.24 and receptor ZO6) are activated by amino acids (Speca et al. 1999; Luu et al. 2004). Further analysis of zebra fish *OlfC* genes has shown that they share eight conserved amino acids, which is a signature motif of other amino acid–sensing ligand-binding receptors (Bertrand et al. 2002; Acher 2005; Alioto and Ngai 2006). In order to fully characterize and understand olfaction in salmonids, it is necessary to first identify all the putatively functional olfactory receptors. It is surprising that there is not more known about the genomics of olfaction in salmonids because of its crucial role in migration and consequently the population structure of wild populations. Three subfamilies of *OlfC* genes have previously been identified in Atlantic salmon, SVRA, SVRB, and SVRC (Dukes et al. 2004, 2006). Here, we report the genome organization of the two *OlfC* gene clusters in Atlantic salmon, the assignment of these genes into subfamilies, and a minimal estimate of the number of functional *OlfC* genes in this species.

## Materials and Methods

### Screening the Atlantic Salmon CHORI-214 BAC Library for *OlfC* Genes

An Atlantic salmon BAC library, CHORI-214 (Thorsen et al. 2005), was obtained from BACPAC Resources, Children's Hospital Oakland Research Institute (CHORI), Oakland, CA. Filters 1–6 were screened by hybridization with the SVRA R (CAGGAGTCCAATATAGCCCAACACAGCCCAGAAGCCAATA) oligonucleotide probe as per the CHORI protocol with the following modifications: Prehybridization was carried out in 5× saline-sodium citrate buffer (SSC), 0.5% sodium dodecyl sulfate (SDS), and 5× Denhardt's solution at 65 °C. The filters were washed three times at 50 °C, 1 h for each wash in 1× SSC and 0.1% SDS. The hybridized filters were placed in phosphor screens overnight and visualized using a Typhoon Imaging system. Hybridization-positive BACs were isolated and grown at 37 °C on a shaker (250 rpm) overnight in 5 ml of Luria-Bertani broth containing chloramphenicol (50 μg/ml). The hybridization-positive BAC DNA was then amplified with the SVAR R primers to verify that these BACs contain the sequence of interest. The polymerase chain reaction (PCR) conditions were: an initial denaturing for 5 min at 95 °C, then 35 cycles of 30 s at 95 °C, annealing temperature of 51 °C for 45 s, extension of 45 s and then a final extension for 5 min at 72 °C, and then stored at 4 °C. PCR products were separated by electrophoresis on a 1.5% agarose gel and visualized with ethidium bromide staining.

### Shotgun Library of the Five *OlfC* Containing BACs

BAC DNA from BACs S0493L23, S0136C02, S0152K21, S0039E15, and S0129H02 was isolated using the Qiagen Large-Construct Kit (Qiagen, Valencia, CA). The isolated DNA from each BAC was sheared by sonication and then blunt-end repaired. The DNA was then size fractioned by agarose gel electrophoresis, the region containing fragments in the 2–5-kb range was excised from the gel, and the DNA was purified using a Qiagen Gel Purification kit. The fragments were ligated into pUC19 plasmid cut with *Sma*I and treated with shrimp alkaline phosphatase to produce dephosphorylated blunt ends, and the ligation mix was used to transform supercompetent *Escherichia coli* cells (Stratagene, La Jolla, CA). Sixty-four recombinant plasmids were digested with *Pvu*II to verify the plasmids contained inserts in the size range 2–5 kb, and then 2,304 clones were sent to the Michael Smith Genome Sciences Centre for sequencing. The sequences were analyzed with Phred/Phrap (Ewing and Green 1998; Ewing et al. 1998) and the results viewed using Consed (Gordon et al. 1998).

### Annotation and Comparative Genomics

The BAC sequences were annotated using the GRASP Annotation Pipeline (grasp.mbb.sfu.ca). ClustalW (Chenna et al. 2003) was used to align the Atlantic salmon OlfC inferred amino acid sequences with those from medaka, three-spined stickleback (*Gasterosteus aculeatus*), zebra fish, green spotted pufferfish, fugu, and the elephant shark (*Callorhinchus milii*) (Alioto and Ngai 2006; Grus and Zhang 2009). Amino acid sequences inferred from another family C GPCRs, V2R2, and two families of taste receptors, T1R 1 and T1R 2 from several teleosts and the elephant shark were used as outgroups. Phylogenetic trees were constructed in Mega4 using the Neighbor-Joining method, and the confidence of each node examined by the bootstrap method with 10,000 replications (Tamura et al. 2007).

### Mapping the OlfC Clusters 1 and 2 Loci in Atlantic Salmon

Polymorphic microsatellite markers, Ssa10050BSFU and Ssa10080BSFU, were identified in the BAC sequences of S0136C02 and S0039E15, respectively. Ssa10050BSFU marker has a 300-bp region containing a (ATCT)55 repeat sequence, and a PCR product was amplified using the following primers: TGTAAAACGACGGCCAGTGACTCCCCACAGAAG and GATGGATGGATGGATGATAG. Ssa10080BSFU marker having a 300-bp region containing a (CA)62 repeat sequence was selected, and a PCR product was amplified using the following primers: TGTAAAACGACGGCCAGTGTTTTCCATCCTGCCTGTCT and TGAGCGTGTTGCGTTCTATG. The first primer for both markers contains an M13 sequence tag that was used in

**Table 2**
**Primers Designed from BAC end Sequences Used to Identify the Minimum Tiling Path across Contigs 859, 1563, and 2358**

| Primer Name | Sequence (5′–3′) |
|---|---|
| S0161P20SP6F | AATGCTCAGTGGTGCTCAAATTA AGCAGTTTGTTACCATA |
| S0161P20SP6R | CTGGAGTTATTGTTGCGACTCT |
| S0161P20T7F | ATTGTGTGCCTTGTGGTTGGAATA CTGTGCTTGAAAAGCA |
| S0161P20T7R | GTGAGGGGAGAGAGTGATGG |
| S0088O06SP6F | TGGGTTTGGTGTCAGTTTGTATCA GTTTGATGTCAGAATG |
| S0088O06SP6F | GTTCCTCTCCATCGCTCGT |
| S0088O06T7F | TCTTCCTCATTGCTGTTGCTAGCAC TTGCCCCCAAAAATT |
| S0088O06T7R | GGTGTGTTTGTGGGGTTAGG |
| S0112J02SP6F | GTCTTTCCCACCGTCCTCTCCTGAC CGTTCATCAGAGCCA |
| S0112J02SP6R | TTCTACCAGGCGAAGCTAATG |
| S0112J02T7F | TGTTGGCTTTCTGCTGTGTTTGAGA AGGCAAAGCCCTTTA |
| S0112J02T7R | GCAGTCCAGCCAAGAACA |

the genotyping analysis. The markers were mapped in the two Atlantic salmon SALMAP mapping families, Br5 and Br6. These two families each contain two parents and 46 offspring (Woram et al. 2003; Danzmann et al. 2005). The genotyping results were analyzed with LINKMFEX software Version 2.3 (Danzmann 2006).

### Qualitative Reverse Transcriptase-PCR of *OlfC* Genes in Atlantic Salmon Olfactory Rosette

A hatchery raised presmolt from the Duncan Hatchery (Duncan, British Columbia) was humanely killed with an overdose of MS222 buffered with $NaHCO_3$. The following tissues were removed: olfactory tissues, brain, heart, head kidney, liver, pyloric caeca, spleen, and muscle and were placed in RNALater. RNA was extracted from each tissue using TRIzol reagent (Invitrogen, Carlsbad, CA). The samples were then transferred to an RNeasy Mini Kit column (Qiagen) and then treated with DNase1 as per the manufacturer's instructions. Total RNA (1 μg) was used to synthesize first-strand cDNA using a Fermentas First-Strand cDNA Synthesis kit and Superscript Reverse Transcriptase (Invitrogen). The cDNA was diluted 10× for use in RT-PCR reactions. Atlantic salmon $\beta$-actin primers were used as a control. OlfC and $\beta$-actin primer sequences are given in the supplementary data S1, Supplementary Material online.

## Results and Discussion

### Identification and Characterization of Atlantic Salmon BACs Containing *OlfC* Genes

The sequences of four salmon *OlfC* genes were aligned using ClustalW (Chenna et al. 2003), and this information was used to design a 40-mer hybridization probe that also served as a reverse PCR primer (SVRA R probe). Only three specific 20-mer PCR primers could be designed for the *OlfC* genes (table 1). The SVRA R probe primer was used as a probe to screen the first six filters of the CHORI-214 Atlantic salmon BAC library, containing 107, 307 BAC clones with an average insert size of 189 kb and giving a 6.8× genome coverage (Thorsen et al. 2005). We identified 36 hybridization-positive BACs, which belong to three contigs (859, 1563, and 2358) based on *Hin*d III fingerprinting (Ng et al. 2005). These BACs were also verified by PCR. BACs from all three contigs were positive for the DQ375533 SVRA. BACs from contig 859 were also positive for DQ375535 SVRA, and BACs from contig 1563 and contig 2358 were positive for DQ375537 SVRB.

BAC minimum tiling paths were constructed for contigs 859, 1563, and 2358. PCR primers were designed from the sequences of SP6 and T7 ends of BACs found in each contig and were used to orient the BACs within each contig (table 2). The BAC end sequence information for the BACs can be found on ASalBase (www.asalbase.org). The minimum tiling paths consisted of one to two BACs for each contig: S0136C02 and S0493L23 for contig 859, S0152K21 and S0039E15 for contig 1563, and S0129H02 for contig 2358 (fig. 1). Contigs 1563 and 2358 were subsequently joined by the marker S0112J02_SP6, which was found in BAC S039E15 of contig 1563 and in BAC S0129H02 in contig 2358. Therefore, contigs 1563 and 2358 became one contig and is now called contig1563. The five minimum tiling path BAC clones from the two contigs were chosen for shotgun sequencing.

### Genomic Organization of the OlfC Loci in Atlantic Salmon

We have identified two clusters of *OlfC* genes in the Atlantic salmon genome. The microsatellite marker Ssa10050BSFU, which was derived from the sequence of BAC S0136C02 from contig 859, could only be mapped in the Atlantic salmon SALMAP mapping family Br5 (Woram et al. 2003; Danzmann et al. 2005) and assigned to Atlantic salmon linkage group 10 (LG10) in the female map. The microsatellite marker Ssa10080BSFU, which was derived from the sequence of BAC S0039E15 from contig

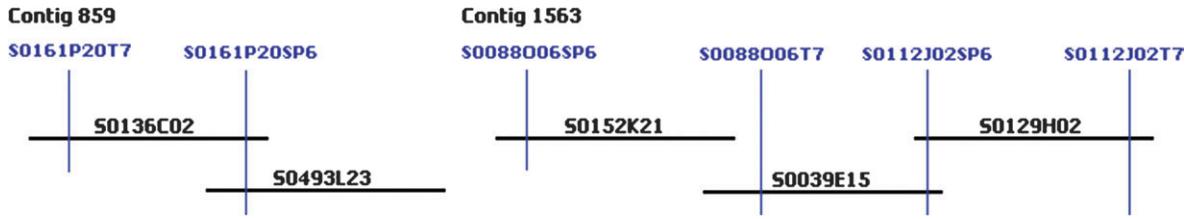


FIG. 1.—Minimum tiling path across the two fingerprint contigs 859 and 1563. The BACs are indicated in black, and the markers used to construct the minimum tiling path are indicated in blue.

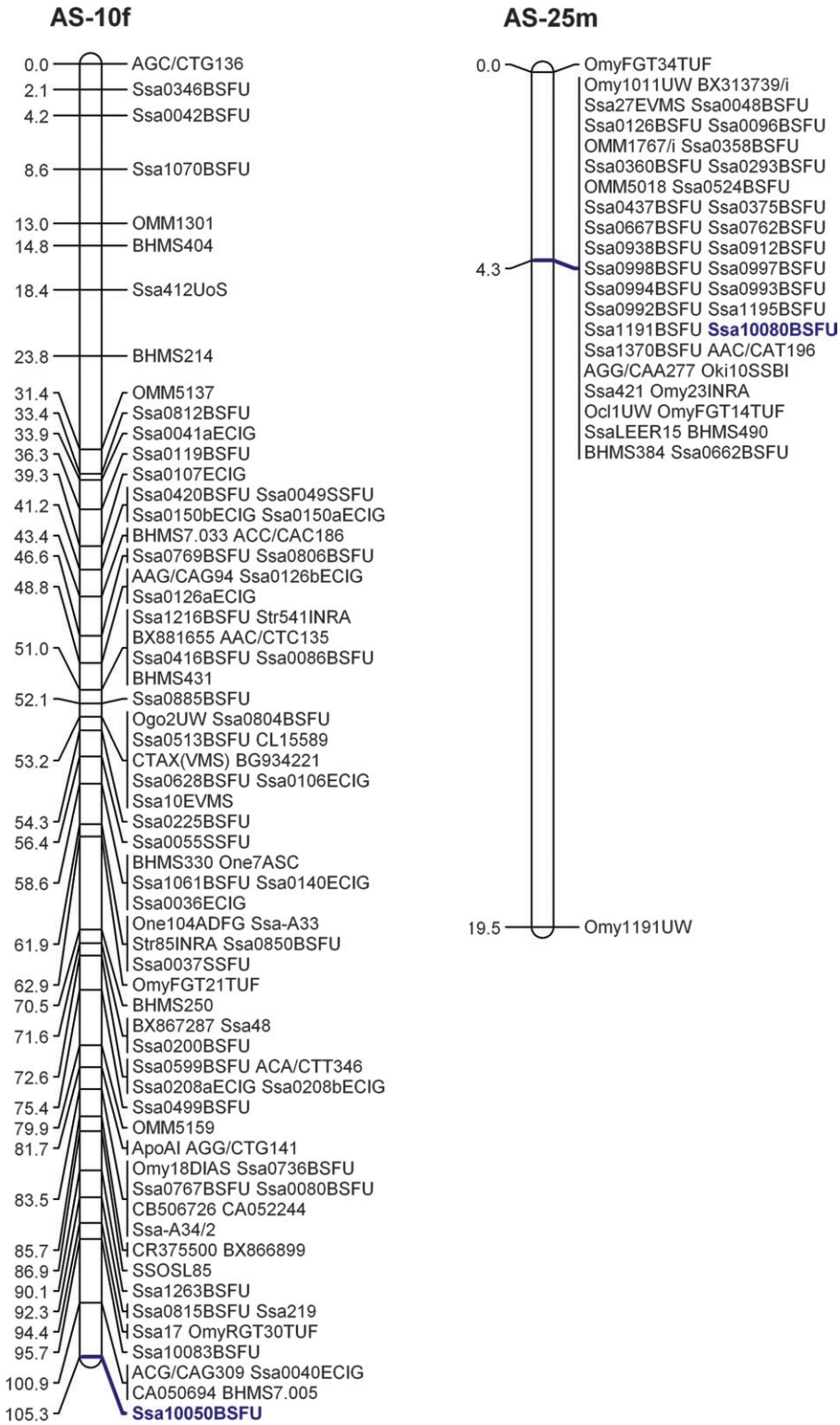


FIG. 2.—Female Linkage group 10 and male linage group 25 (AS-10f, AS-25m) of the Atlantic salmon mapping family Br5 and Br6, respectively, showing the location of Ssa10050BSFU and Ssa10080BSFU. These microsatellite markers were designed from the sequences of BACs S0136C02 and S0039E15, respectively.

1563, could only be mapped in the Br6 mapping family and was assigned to Atlantic salmon linkage group 25 (LG25) in the male map (fig. 2). Fluorescent *in situ* hybridization analysis validated the results of the linkage analysis. BAC S0136C02 hybridized to the telomeric end of a single chromosome pair, chromosome 9 (LG10), and BAC S0152K21 hybridized to the telomeric end of chromosome 20 (LG25) (results not shown).

**Table 3**
**BAC Clone Sequencing Results**

| Name of BAC | GenBank Accession Number | Size of the BAC (bp) | Initial No. of Individual Contigs | Initial No. of Scaffolds | No. of Individual Contigs after Finishing | No. of Scaffolds after Finishing |
|---|---|---|---|---|---|---|
| S0136C02 | FJ423038 | 195,844 | 9 | 3 | 4 | 4 |
| S0493L23 | FJ423033 | 248,385 | 19 | 4 | 8 | 4 |
| S0152K21 | FJ423035 | 183,554 | 19 | 5 | 14 | 4 |
| S0039E15 | FJ423037 | 206,868 | 14 | 4 | 11 | 7 |
| S0129H02 | FJ423036 | 214,342 | 8 | 2 | 5 | 2 |

Contigs are defined as any sequences over 1,500 bp and containing more than 10 reads; scaffolds are defined as any sequences over 10,000 bp.

## Annotation of the Sequence of the Five OlfC Containing BACs

The shotgun libraries of the BAC clones were sequenced at the Michael Smith Genome Sciences Centre, Vancouver, and annotated using the GRASP annotation pipeline (grasp.mbb.sfu.ca). Table 3 shows the results of the initial assemblies of the shotgun sequences. Contig orientation and gap closing were carried out by designing primers from the ends of the contigs, amplifying the gap regions, and subsequently sequencing the gap amplicons. Due to the presence of repetitive sequences, not all of the gaps could be sequenced, but the orientation and most of the gap sizes were estimated. The sequences have been deposited in GenBank (accession nos. FJ423033, FJ423035–FJ423038). The BAC sequences were then assembled into two sequence scaffolds 859 (BACs S0136C02 and S0493L23) and 1563 (S0152K21, S0039E15, and S0129H02) using Phred/Phrap and manual inspection (www.asalbase.org).

We identified 55 putative *OlfC* genes in the Atlantic salmon genome. *OlfC* genes were classified as putatively functional if they had a complete coding sequence, the six stereotypical OlfC exons, and seven predicted transmembrane domains, which are predicted to be of a certain length required to span the membrane. Genes that failed any of these criteria were considered pseudogenes. Using these criteria, we identified 29 putatively functional genes and 26 pseudogenes (table 4, supplementary data S2 and S3, Supplementary Material online). This prediction of the *OlfC* genes in Atlantic salmon is a minimal estimate of the true number of genes because the sequenced BAC sequences are gapped; however, this will be verified when the sequence of the Atlantic salmon genome becomes available by the middle of 2010.

The OlfC family may function as amino acid–sensing receptors in fish. It has been shown that two orthologous receptors from goldfish and zebra fish are activated by amino acids (Speca et al. 1999; Luu et al. 2004). The N-terminal ligand-binding domain (NTD) of family C GPCRs is thought to form a bilobate clamshell-like conformation (Parmentier et al. 2002; Pin et al. 2004). The binding pocket has two lobes, the proximal pocket and the distal pocket (Acher and Bertrand 2005). Several studies on amino acid binding receptors have identified eight signature residues in the NTD (Bertrand et al. 2002; Acher and Bertrand 2005; Alioto and Ngai 2006), with five conserved residues in the proximal binding pocket, which is used in ligand binding, and three residues in the hinge region. Mutagenesis of any of the five residues in the proximal binding pocket of an orthologous *OlfC* gene in goldfish (5.24) yields a decrease in receptor binding activity with the ligand (Luu et al. 2004). These eight signature motif residues are necessary for proper ligand binding and should be conserved in all amino acid binding proteins. These eight signature residues are conserved in 24 of the 29 Atlantic salmon OlfC receptors, suggesting that these receptors also function as amino acid binding proteins (supplementary data, S4, Supplementary Material online).

## Phylogenetic Analysis of Putatively Functional *OlfC* Genes

Phylogenetic analysis of the inferred amino acid sequences of all putatively functional *OlfC* genes identified

**Table 4**
**Number of *OlfC* Genes Belonging to Different Subfamilies in Five Fish Species and the Elephant Shark**

| | Species and Abbreviations | | | | | |
|---|---|---|---|---|---|---|
| Subfamily[a] | Atlantic Salmon Ssa | Zebra Fish[b] Dre | Medaka[b] Ola | Fugu[b] Tru | Green Spotted Pufferfish[b] Tni | Elephant Shark[c] Cmi |
| 1 | 0 (0) | 1 (0) | 0 (0) | 0 (0) | 0 (0) | — |
| 2 | 2 (4) | 1 (2) | 1 (0) | 1 (0) | 1 (0) | — |
| 3 | 1 (0) | 1 (0) | 1 (0) | 0 (0) | 0 (0) | — |
| 4 | 11 (4) | 3 (0) | 6 (8) | 5 (0) | 1 (4) | — |
| 5 | 0 (0) | 1 (0) | 0 (0) | 0 (0) | 0 (0) | — |
| 6 | 0 (0) | 0 (0) | 0 (0) | 2 (0) | 2 (0) | — |
| 7 | 0 (0) | 1 (0) | 1 (0) | 1 (0) | 1 (0) | — |
| 8 | 0 (0) | 0 (1) | 2 (2) | 3 (8) | 2 (0) | — |
| 9 | 0 (0) | 13 (2) | 1 (0) | 2 (2) | 1 (1) | — |
| 10 | 0 (0) | 1 (1) | 2 (4) | 1 (0) | 1 (0) | — |
| 11 | 5 (1) | 1 (0) | 0 (0) | 0 (0) | 0 (0) | — |
| 12 | 1 (0) | 1 (2) | 1 (0) | 1 (0) | 1 (0) | — |
| 13 | 1 (0) | 1 (0) | 1 (0) | 0 (0) | 0 (0) | — |
| 14 | 1 (0) | 1 (0) | 0 (2) | 1 (0) | 0 (1) | — |
| 15 | 2 (0) | 0 (0) | 1 (0) | 4 (0) | 1 (1) | — |
| 16 | 3 (1) | 19 (1) | 2 (3) | 6 (2) | 0 (4) | — |
| 17 | 2 (3) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | — |
| Other* | (13) | — | — | — | — | — |
| A | — | — | — | — | — | 11 |
| B | — | — | — | — | — | 3 |
| Total | 29 (26) | 45 (9) | 17 (19) | 27 (12) | 11 (11) | 14 |

Pseudogenes indicated in the brackets and the other* indicates pseudogenes that do not belong to any of the subfamilies.

[a] Subfamily naming is based on the phylogenetic tree in figure 3.

[b] The deduced amino acid sequences were obtained from Hashiguchi and Nishida (2006).

[c] The deduced amino acid sequences were obtained from Grus and Zhang (2009).

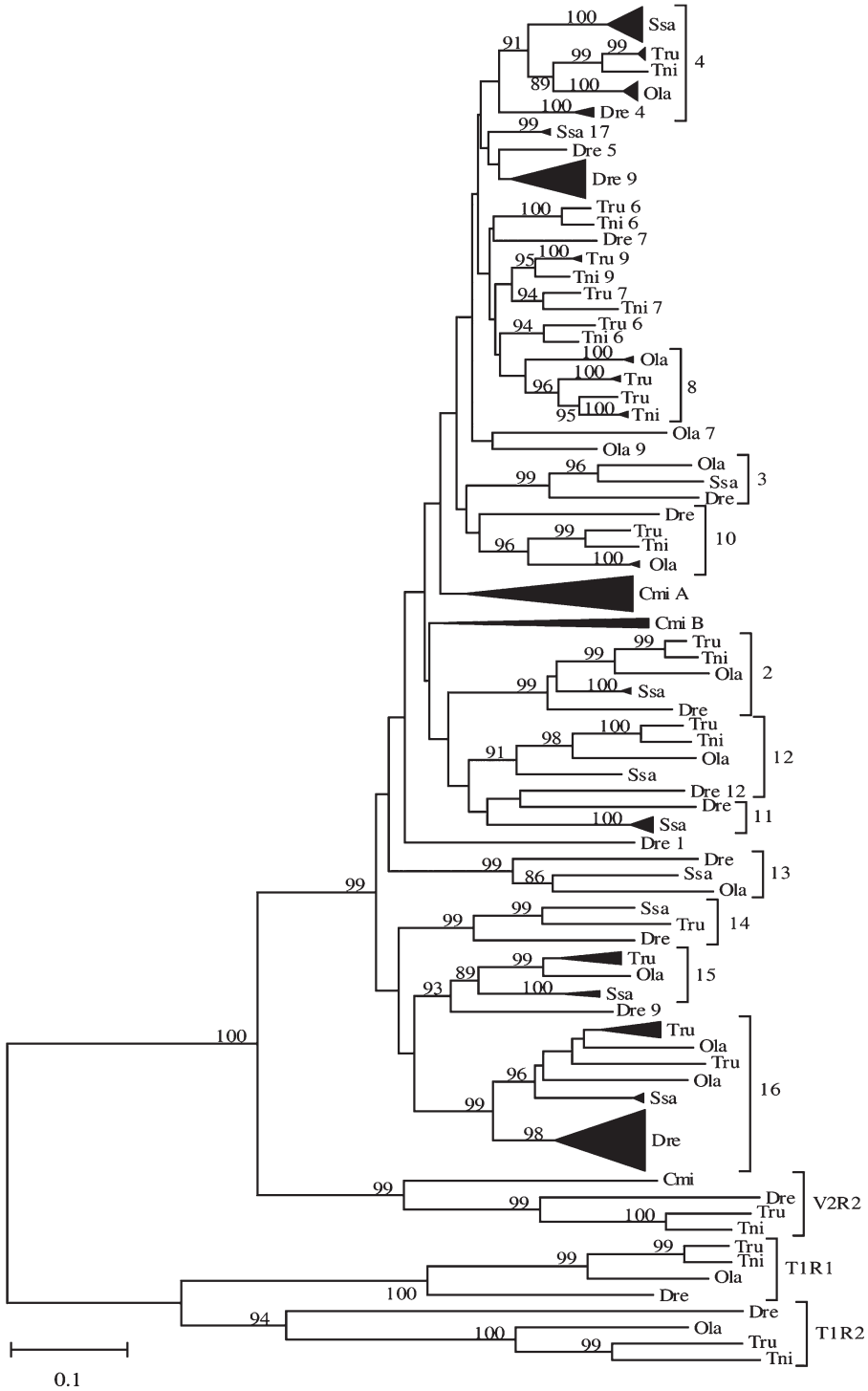


FIG. 3.—Neighbor-Joining phylogenetic tree of 155 taxa with bootstrap support (10,000 replicates) of the amino acid sequence of the teleost *OlfC* functional genes and the elephant shark V2R partial gene sequences. V2R2, T1R 1 and T1R 2 sequences were used as outgroups. The optimal tree has the sum of the branch lengths of 13.03646339. The data set contained 198 positions. The subfamilies are numbered as assigned in the zebra fish *OlfC* genes. An expanded gene family has been indicated by a triangle, and the vertical height of the end of the triangle is proportional to the number of genes in this family, refer to table 4 and/or supplementary data, S5, Supplementary Material online.

in teleosts and the elephant shark reveals that the Atlantic salmon *OlfC* genes can be grouped into 10 subfamilies (table 4 and fig. 3). Seven *OlfC* genes identified as functional genes in other fish were omitted from the analysis because they did not meet the criteria for a functional gene in this study. In silico translation of these genes indicated that the protein products were missing part of a transmembrane region. The elephant shark inferred amino acid sequences

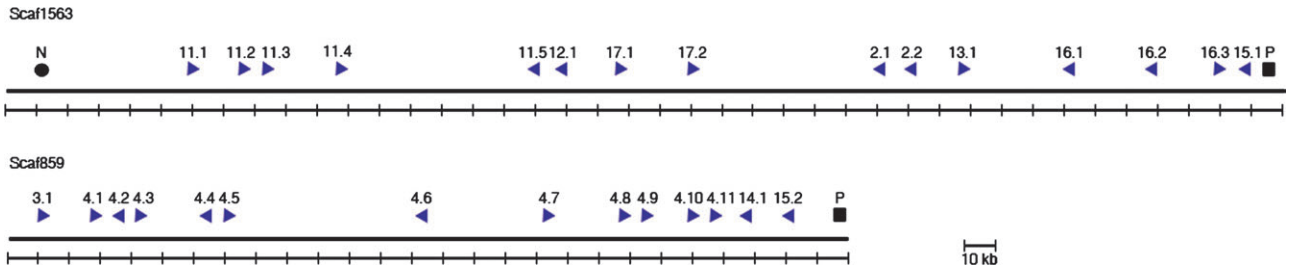


FIG. 4.—Transcriptional orientation of the putatively functional *OlfC* genes in Atlantic salmon. The *OlfC* genes are numbered according to their subfamily and their order in the scaffold. The genes flanking the OlfC clusters are also shown, neprilysin (N) and phospholipase C (P).

represented only partial sequences and were used as an interesting comparison to see if the elephant shark, another aquatic organism, shared similar OlfC receptor genes. However, all the *OlfC* genes that have been identified in the elephant shark form two subfamilies that are distinct from those of the teleosts. Nine of the Atlantic salmon subfamilies correspond to subfamilies found in other teleosts, and one subfamily appears to be salmon specific. The subfamilies have been named according to the convention used for the zebra fish *OlfC* genes (Alioto and Ngai 2006). The naming of the Atlantic salmon *OlfC* genes is based on their subfamily and their physical location in the cluster, and the salmon-specific subfamily (i.e., subfamily 17) was added to this scheme (fig. 4). There is also a large expansion in the number of *OlfC* genes in one subfamily in Atlantic salmon (subfamily 4), as has been observed in other teleosts (table 4). These differences in gene number may reflect species-specific evolutionary requirements for olfaction. It will be interesting to determine if the same families are amplified in all salmonid species and also in different populations within a species.

The Atlantic salmon pseudogenes were also analyzed with phylogenetic analysis to determine which subfamily they had evolved from. Half of the pseudogenes could be placed within a subfamily, and the other half were too divergent to assign them to a subfamily (table 4).

## *OlfC* Gene Expression

The qualitative Reverse Transcriptase-PCR results revealed that all of the putative functional *OlfC* genes that were tested were only expressed in the olfactory rosette of a presmolt Atlantic salmon. This is consistent with these genes encoding biologically active OlfC ORs (supplementary data, S1, Supplementary Material online).

## Evolution of *OlfC* Gene Clusters

After the tetrapod–teleost divergence, it appears that the common ancestor of the teleosts had an *OlfC* gene repertoire containing many of the subfamilies of *OlfC* genes found in extant fish. The cluster of *OlfC* genes grew through a series of tandem duplications. As fish evolved, there were further tandem duplications and losses of *OlfC* genes in different lineages (Hashiguchi and Nishida 2006). Within the Acanthoptergii (e.g., pufferfish, medaka, and three-spined stickleback), a single cluster of *OlfC* genes has been retained. However, within the genomes of zebra fish and Atlantic salmon, two clusters have been identified.

The Atlantic salmon *OlfC* gene cluster on chromosome 9 is flanked by two genes, neprilysin and the $\eta$-type phospholipase C (PLC) (fig. 4). These two genes have been identified as flanking genes to the major cluster of *OlfC* genes in other teleosts except zebra fish, which has only retained the PLC flanking gene in the *OlfC* gene clusters on chromosome 17 and 18 (Hashiguchi and Nishida 2006). The other salmon cluster on chromosome 20 is also flanked by PLC but has not retained the flanking gene, neprilysin, at the other end. It appears that one of the salmon *OlfC* gene clusters is similar to the genomic structure found in medaka, stickleback, and fugu, whereas the other resembles the zebra fish structure (fig. 5). These flanking genes are not found in tetrapods; therefore, they must have originated after the divergence of the teleost and tetrapod lineages (Hashiguchi and Nishida 2006). One possibility to account for the formation of these second clusters and the common phospholipase flanking gene order in zebra fish and Atlantic salmon is that they are the result of segmental duplication in the common ancestor to the Ostariophysi (zebra fish) and Protacanthopterygii (Atlantic salmon). Alternatively, the second cluster could have arisen from independent duplication events with the salmon duplicates coming from the whole genome auto-tetraploidization event that occurred in the common ancestor of extant salmonids (Allendorf and Thorgaard 1984).

## Functional Significance of OlfC Cluster(s) Flanking Genes

Mammalian V2Rs have been shown to be stimulated by major histocompatibility complex (MHC) class I peptides (Leinders-Zufall et al. 2004). Neprilysin is a membrane-bound neutral peptidase, which is upregulated during ovulation. It has been hypothesized that peptides may be cleaved by neprilysin and released into the water by individuals or from spawned eggs and perceived by OlfC receptors. These peptides may be used as a means of signaling reproduction and/or genetic identity between individuals, similar to the MHC peptides in mammals and in fish (Leinders-Zufall et al. 2004; Milinski et al. 2005). PLC is involved in the inositol 1,4,5-triphosphate (InsP3) second messenger pathway in the signal transduction of the olfactory receptor cell (Ache and Zhainazarov 1995). This type of PLC has been localized to neurons and other family 3 GPCR members have been shown to be coupled with PLC (Pin et al. 2003). Further

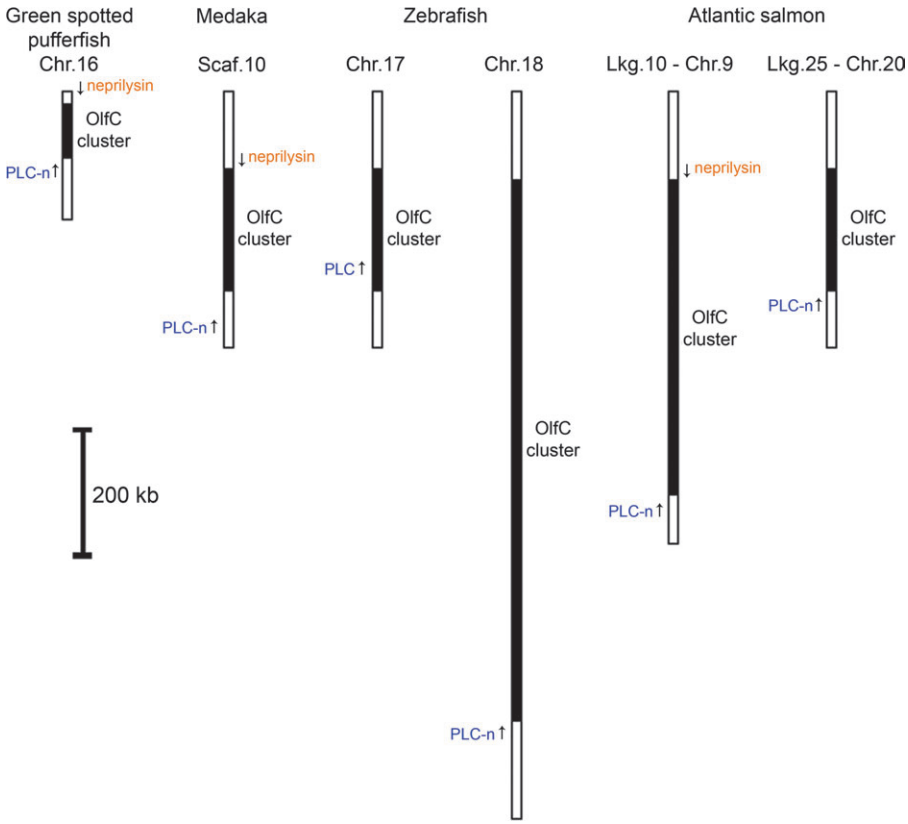


FIG. 5.—Chromosomal location and conserved synteny of flanking genes surrounding the major *OlfC* clusters in gereen spotted pufferfish, medaka, zebrafish, and the two *OlfC* clusters in the Atlantic salmon genome. In all teleost examined, the *OlfC* clusters are flanked by PLC, and for the majority, the other end of the cluster is flanked by neprilysin (figure adapted from Hashiguchi and Nishida 2006). PLC denoted on chromosome 17 of zebra fish contains a phospholipase C domain.

evidence has shown that odorants activate PLC in fish olfactory receptor neurons (Lo et al. 1994; Ache and Zhainazarov 1995). This suggests that the PLC may have been maintained close to the *OlfC* gene cluster(s) in all teleosts because of its functional importance in the signal transduction in the olfactory cell.

## Conclusions

We have characterized a minimal estimate of the *OlfC* gene repertoire of Atlantic salmon. We identified 29 putatively functional genes and 26 pseudogenes. Atlantic salmon OlfC subfamily 17 does not occur in the other fish genomes that have been examined to date. In addition, there appears to have been an expansion of subfamily 4 *OlfC* genes. The members of these subfamilies are particularly worthy of further study to determine if they occur and/or have expanded in other salmonids and if they are involved in the migration and homing response that is characteristic of these species.

## Supplementary Material

Supplementary data S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

## Acknowledgments

We thank the Sequencing Team at the Michael Smith Genome Sciences Centre for sequencing the BAC shotgun libraries. This research was supported by Genome Canada, Genome BC, and NSERC (post-graduate scholarship to K.A.J.).

## Literature Cited

Ache BW, Zhainazarov A. 1995. Dual second-messenger pathways in olfactory transduction. Curr Opin Neuro Biol. 5:461–466.

Acher FC, Bertrand HO. 2005. Amino acid recognition by Venus flytrap domains is encoded in an 8-residue motif. Biopolymers. 80:357–366.

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

Alioto TS, Ngai J. 2006. The repertoire of olfactory C family G protein-coupled receptors in zebrafish: candidate chemosensory receptors for amino acids. BMC Genomics. 7:309.

Allendorf F, Thorgaard GH. 1984. Evolutionary genetics of fishes. New York: plenum Press.

Bertrand HO, Bessis AS, Pin JP, Acher FC. 2002. Common and selective molecular determinants involved in metabotropic glutamate receptor agonist activity. J Med Chem. 45:3171–3183.

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31:3497–3500.

Danzmann RG. 2006. LINKMFEX: linkage analysis package for outcrossed families with male or female exchange of the mapping parent version 2.3.

Danzmann RG, Cairney M, Davidson WS, et al. 2005. A comparative analysis of the rainbow trout genome with 2 other species of fish (Arctic charr and Atlantic salmon) within the tetraploid derivative Salmonidae family (subfamily: salmoninae). Genome. 48:1037–1051 (14 co-authors).

Dukes JP, Deaville R, Bruford MW, Youngson AF, Jordan WC. 2004. Odorant receptor gene expression changes during the parr–smolt transformation in Atlantic salmon. Mol Ecol. 13:2851–2857.

Dukes JP, Deaville R, Gottelli D, Neigel JE, Bruford MW, Jordan WC. 2006. Isolation and characterisation of main olfactory and vomeronasal receptor gene families from the Atlantic salmon (*Salmo salar*). Gene. 371:257–267.

Ewing B, Green P. 1998. Basecalling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8:186–194.

Ewing B, Hillier L, Wendl M, Green P. 1998. Basecalling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8:175–185.

Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. Genome Res. 8:195–202.

Grus WE, Zhang J. 2009. Origin of the genetic components of the vomeronasal system in the common ancestor of all extant vertebrates. Mol Biol Evol. 26(2):407–419.

Han G, Hampson DR. 1999. Ligand binding to the amino-terminal domain of the mGluR4 subtype of metabotropic glutamate receptor. J Biol Chem. 274:10008–10013.

Hashiguchi Y, Nishida M. 2006. Evolution and origin of vomeronasal-type odorant receptor gene repertoire in fishes. BMC Evol Biol. 6:76.

Johnstone KA, Lubieniecki KP, Chow W, Phillips RB, Koop BF, Davidson WS. 2008. Genomic organization and characterization of two vomeronasal 1 receptor-like genes (ora1 and ora2) in Atlantic salmon *Salmo salar*. Mar Genomics. 1:23–31.

King TL, Kalinowski ST, Schill WB, Spidle AP, Lubinski BA. 2001. Population structure of Atlantic salmon (*Salmo salar L.*): a range-wide perspective from microsatellite DNA variation. Mol Ecol. 10:807–821.

Leinders-Zufall T, Brennan P, Widmayer PSPC, Maul-Pavicic A, Jager M, Li XH, Breer H, Zufall F, Boehm T. 2004. MHC class I peptides as chemosensory signals in the vomeronasal organ. Science. 306:1033–1037.

Lo YH, Bellis SL, Cheng LJ, Pang J, Bradley TM, Rhoads DE. 1994. Signal transduction for taurocholic acid in the olfactory system of Atlantic salmon. Chem Sens. 19:371–380.

Luu P, Acher F, Bertrand HO, Fan J, Ngai J. 2004. Molecular determinants of ligand selectivity in a vertebrate odorant receptor. J Neurosci. 24:10128–10137.

Milinski M, Griffiths S, Wegner KM, Reusch TB, Haas-Assenbaum A, Boehm T. 2005. Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. Proc Natl Acad Sci USA. 102:4414–4418.

Ng SHS, Artieri CG, Bosdet IE, et al. 2005. A physical map of the genome of Atlantic Salmon, *Salmo salar*. Genomics. 86:396–404 (29 co-authors).

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

Parmentier ML, Prezeau L, Bockaert J, Pin JP. 2002. A model for the function of family 3 GPCRs. Trends Pharmacol Sci. 23: 268–274.

Pin JP, Galvez T, Prezeau L. 2003. Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors. Pharmacol Ther. 98:325–354.

Pin JP, Kniazeff J, Goudet C, Bessis AS, Liu J, Galvez T, Archer F, Rondard P, Prezeau L. 2004. The activation mechanism of class-C G-protein coupled receptors. Biol Cell. 96:335–342.

Saraiva LR, Korsching SI. 2007. A novel olfactory receptor gene family in teleost fish. Genome Res. 17:1448–1457.

Shi P, Zhang J. 2007. Comparative genomic analysis identifies an evolutionary shift of vomeronasal receptor gene repertoires in the vertebrate transition from water to land. Genome Res. 17: 166–174.

Speca DJ, Lin DM, Sorenson PW, Isacoff EY, Ngai J, Dittman Ah. 1999. Functional identification of a goldfish odorant receptor. Neuron. 23(3):487–498.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol. 24:1596–1599.

Thorsen J, Zhu B, Frengen E, Osoegawa K, de Jong PJ, Koop BF, Davidson WS, Hoyheim B. 2005. A highly redundant BAC library of Atlantic salmon (*Salmo salar*): an important tool for salmon projects. BMC Genomics. 6:50.

Woram RA, Gharbi K, Sakamoto T, et al. 2003. Comparative genome analysis of the primary sex-determining locus in salmonid fishes. Genome Res. 13:272–280 (16 co-authors).

David Irwin, Associate Editor

Accepted February 6, 2009