A NOVEL OLFACTORY RECEPTOR GENE FAMILY IN TELEOST FISH: PHYLOGENOMICS, CELLULAR LOCALIZATION AND COMPARISON WITH OTHER OLFACTORY RECEPTOR GENE FAMILIES

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Luis Miguel Rodrigues Saraiva

aus Manteigas, Portugal

Köln 2008

Berichterstatter: Prof. Dr. S. I. Korsching

Prof. Dr. Jens Brüning

Tag der mündlichen Prüfung: 30. Juni 2008

ERKLÄRUNG

Ich versichere, dass ich die von mir vorgelegte Dissertation selbstständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen in der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfungvorgelegen hat; dass sie - abgesehen von den unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist, sowie dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. S. I. Korsching betreut worden.

Köln, den 5. Mai, 2008

Teilpublikationen:

- Hussain, A*; **Saraiva, LR*** & Korsching SI (**2008**). "*Strong positive selection and high intron dynamics in teleost taar genes*". PNAS *under revision.* *both authors contributed equally to this paper

- Kraemer, AM; **Saraiva, LR** & Korsching, SI (**2008**). "*Structural and functional diversification in the teleost S100 family of calcium-binding proteins*". BMC Evolutionary Biology, 8:48

- **Saraiva, LR** & Korsching, SI (**2007**). *"A novel olfactory receptor gene family in teleost fish".* Genome Research. 17:1448-57

„Regozijarme-ei apenas com a diversidade da natureza

e deixarei os problemas concretos

para os políticos e pregadores”

Stephen Jay Gould

PARA OS MEUS PAIS,

I. ACKNOWLEDGEMENTS

This study was performed in the laboratory of Prof. Dr. Sigrun Korsching in the Institute of Genetics, Cologne, Germany and was supported by a grant from the International Graduate School of Genetics and Functional Genomics (IGS-GFG). I would like to start this section by thanking her first for accepting being my mentor, second for being my "scientific mother" and third for becoming a very good friend. All the encouragement, support, advice and everything else throughout these three and a half years are priceless. Her patience to deal with my "awkwardness" is also treasured ☺

I would also like to acknowledge to Prof. Dr. Jens Brüning and to Prof. Dr. Peter Kloppenburg for accepting to be in my thesis committee, for the interest in my work, nice feedback and for accepting serving as my referees.

Being accepted in the IGS-GFG and to be part of the Class of 2004 made first of all my presence is Germany possible and gave me the chance to learn a lot more outside of the scientific environment. The cultural exchange was and is a great thing and also helped me becoming a better person and scientist. In this context I would like to give a special "Thank You" to Brigitte Wilcken-Bergmann for all the help and support throughout all these years. Thank you again.

Next in line are Leo Kurian, Dale Richardson and Lars Rosvoll (Drs yet to be) for their valuable friendship, motivation, inspiration, patience, company, tripod or bipod moments in India, the Dam, Istambul, Bratislava, Australia, for being my climbing and/or football partners… and for all the other things that would not be wise to mention here. Words are not enough to describe the friendship built during these years.

Also to Yen Yen and Jinu, the "chicks" of the group for all the support, cooking, good and bad moments shared and everything that it is yet to come. Yen

To my lab mates, especially Dr Yuichiro Oka for valuable feedback, help and collaboration in several projects. Also, Andreas Kraemer and Ashiq Hussain and again Yen Yen should be mentioned in this context for the fruitful collaboration and by providing me some of the figures used in this thesis. To Mehmet, for all the happiness and joy and of course, for being the step father of all our fish. To all of you a very big thank you for the learning experience and moments shared until now.

To all the others that I have met since I arrived in Germany. Just to mention a few: Channa K.C., Cemali, Daniella, Joanna, Rodrigo, Natália, Henrique, Baghi (I am sure I misspelled his name again), Raja, the 4040 crew, Gesa… and all the others.

Also to my very good friends from Portugal: Bruno, João, Diogo, Nuno, Neiva, Graf, Sérgio, Vazé, Yana and the rest of Évora Crew. Àos "Caravela" pelos excelentes momentos passados, apoio e por me ajudarem a matar a saudade de Portugal. To my old colleagues from the IGC for broadening my scientific horizons before I even applied to Cologne.

Last but not least to my family, for everything. They are the main reason why I am here.

Obrigado Pai, Mãe, Marta e Rafa. Mesmo estando fisicamente cada vez mais longe, não me esqueço nem nunca esquecerei de vós. Vocês estão sempre bem junto a mim. Não há palavras para descrever todo o apoio ao longo destes vinte e quase oito anos. Um obrigado cheio de maiúsculas. Esta tese é vossa também.

II. CONTENTS

III. FIGURE LIST

IV. ABSTRACT

While for two of four mammalian olfactory receptor families, all of them G protein-coupled receptors, ortholog teleost families have been identified and well-characterized (OR and V2R), two other families (V1R and TAAR) lack to date a systematic study in non-mammalian vertebrates.

By data mining I identified a total of six V1R-like genes in five teleost species plus four orthologs in one jawless and one cartilaginous fish species each. In the phylogenetic analysis these *ora* genes (olfactory receptor, class A-related) form a single clade with three subclades, one of them including the entire mammalian V1R superfamily. The Ora family originates early in the vertebrate lineage, before the separation of the jawless from jawed fish. A similar search was performed also for *taar* genes in genomes of five teleosts, two basal fish and seven higher vertebrates. *Taar* genes segregate into three classes and their family size ranges from 18 to 112 genes in teleosts (pufferfish and zebrafish, respectively), while mammalian families contain at most 19 genes (opossum*)*. The TAAR family originated in the common ancestor of bony and cartilaginous fishes, after its divergence from jawless fish. In these and other properties the *ora* and *taar* gene families turn out to be at opposite poles of the spectrum of olfactory receptor families.

All the six teleost Ora family members are evolutionarily much older than the speciation events in the teleost lineage, while most extant teleost *taar* genes have emerged late in evolution, well after the split between basal teleosts (zebrafish) and *neoteleostei* (stickleback, medaka, pufferfish). *Taar* genes are largely arranged according to phylogenetic proximity in two big clusters (both syntenic to the single sarcopterygian gene cluster), whereas the *ora* genes are organized as singletons or symmetrical gene pairs. *TAAR* genes are mostly monoexonic, whereas two *ora* genes exhibit a highly conserved multi-exonic structure. Furthermore, the *ora* genes are under strong negative selection (minute *d*N/*d*S values), whereas the teleost taar genes display a relaxed pattern of global negative selection and an unprecedented degree of local positive selection.

Taken together, the *ora* gene repertoire is highly conserved across teleosts, in striking contrast to the frequent species-specific expansions observed in mammalian V1Rs. The inverse is observed for the taar gene repertoire, which is rather conserved across mammalian species, but exhibits frequent and large species-specific expansions in teleosts. Thus, the transition from teleosts to tetrapods may parallel a transition in function as well as regulation of both the ora/V1R and TAAR gene families.

Consistent with a function as olfactory receptors all zebrafish *ora* and all analyzed *taar* genes (except *taar*1) were expressed in sparse subsets of olfactory receptor neurons. The olfactory epithelium contains three subtypes of olfactory receptor neurons, ciliated, microvillous and crypt cells, the latter so far without known receptors, but with both cilia and microvilli. I found the *ora* genes to be expressed in the crypt cells, thereby deorphanizing this third type of OSN. Furthermore, the *ora* genes follow the monogenic rule of expression previously reported for members of other olfactory receptor gene families. Ora genes co-express both Gαi and Gαo, supporting the hypothesis that crypt cells might possess two distinct olfactory signaling pathways, one via their cilia and the other via microvilli.

V. ZUSAMMENFASSUNG

Nur zwei der vier grossen Familien von G Protein-gekoppelten Geruchsrezeptorgenen sind systematisch in Fischen untersucht worden (OR und V2R), während die restlichen zwei Familien (V1R und TAAR) bisher nur in Säugern genauer analysiert wurden.

Mittels ausgedehnter Datenbankanalysen konnte ich insgesamt sechs V1R-ähnliche Gene in fünf Arten von Knochenfischen identifizieren, sowie je vier Orthologe in Knorpelfischen und kieferlosen Fischen. In der phylogenetischen Analyse bilden diese *ora* Gene (olfaktorische Rezeptoren, verwandt mit Klasse A der G Protein-gekoppelten Rezeptoren) einen einzelnen Klade, der sich aus drei Unterkladen zusammensetzt, von denen einer die gesamte V1R Superfamilie der Säuger enthält. Die *ora* Genfamilie entstand bereits früh in der evolutionären Geschichte der Wirbeltiere, vor der Aufspaltung von kieferlosen Fischen (*Agnatha*) und *Gnathostomata*. Ich habe entsprechende Datenbanksuchen auch für die *taar* Genfamilie in den Genomen von fünf Arten von Knochenfischen, zwei Arten basaler Fische, und sieben Arten landlebender Vertebraten durchgeführt. *Taar* Gene lassen sich in drei Untergruppen/Klassen aufteilen, und pro Art werden zwischen 18 und 112 Genen beobachtet (Kugelfische bzw. Zebrabärbling, *Danio rerio*), während in Säugern maximal 19 Gene gefunden wurden (Opossum). Die *taar* Genfamilie entstand im gemeinsamen Vorläufer der Knochen- und Knorpelfische, nach dessen Abspaltung von kieferlosen Fischen. Hier wie auch in anderen Analysen zeigt sich, dass *ora* und *taar* Genfamilien sehr verschiedene Wege in der Evolution gegangen sind, ja geradezu an den entgegengesetzten Enden des Spektrums olfaktorischer Rezeptorgenfamilien liegen.

Alle sechs *ora* Gene sind lange vor der Speziation der Knochenfische entstanden, während fast alle *taar* Gene der Knochenfische viel später evolvierten, sogar nach der Aufteilung in basale Teleosten (Zebrabärbling) und Neoteleosten (Stichling, Medaka, Kugelfisch). *Taar* Gene sind großenteils gemäß ihrer phylogenetischen Verwandtschaft auf dem Genom angeordnet, in zwei großen Clustern, die beide syntenisch zu dem Cluster der *taar* Gene in landlebenden Vertebraten sind, während die *ora* Gene einzeln bzw. als zwei symmetrische Genpaare vorliegen. Die meisten *taar* Gene bestehen nur aus einem Exon, während zwei der sechs *ora* Gene eine hochkonservierte multi-exonische Struktur aufweisen. *Ora* Gene unterliegen einer starken negativen Selektion, mit extrem kleinen dN/dS Werten, während die *taar* Gene der Knochenfische nur eine moderate globale negative Selektion aufweisen, aber gleichzeitig eine stark ausgeprägte positive Selektion an einzelnen Sequenzpositionen zeigen.

Zusammengefasst läßt sich sagen, dass das *ora* Genrepertoire sehr stabil und hochkonserviert ist, ganz im Gegensatz übrigens zu den häufigen spezies-spezifischen Expansionen der daraus hervorgegangenen Säuger V1R Gene. Umgekehrte Verhältnisse gelten für die *taar* Gene, die innerhalb der Säuger recht gut konserviert sind, aber in den Knochenfischen häufige und spezies-spezifische Genexpansionen aufweisen. Der Übergang von Teleosten zu Tetrapoden könnte daher in beiden Genfamilien mit einem drastischen Wechsel sowohl in Funktion und Genregulation einhergehen.

Wie für Geruchsrezeptoren erwartet, wurden alle *ora* Gene und alle untersuchten *taar* Gene des Zebrabärblings (außer *taar1*) in vereinzelten olfaktorischen Rezeptorneuronen exprimiert. Es existieren drei Arten dieser Neurone, die ziliierten, mikrovillären Neurone und die Kryptzellen, die sowohl Zilien als auch Mikrovilli tragen, jedoch bisher keinen Geruchsrezeptoren zugeordnet werden konnten. Ich konnte zeigen, dass die *ora* Gene spezifisch in Kryptzellen exprimiert werden, und damit die bisher völlig offene Frage einer Funktion dieser Kryptzellen einer Lösung näherbringen. Weiters konnte ich nachweisen, dass die *ora* Gene der Ein-Rezeptorneuron-ein-Rezeptorgen Regel folgen, wie es für Mitglieder verschiedener anderer olfakorischer Rezeptorgenfamilien bekannt ist. *Ora* Gene werden zusammen mit Gαi und Gαo exprimiert, was die Hypothese nahelegt, dass Kryptzellen zwei unabhängige Signalwege aufweisen könnten, einen in den Zilien, und den anderen in den Mikrovilli.

VI. INTRODUCTION

"Nothing in Biology Makes Sense Except in the Light of Evolution"

T. Dobzhansky

How many times has a given smell redirected us to a certain episode of our childhood? In humans, smell is often regarded as the sense that elicits the strongest memories. Thus, the addition of emotional qualities to previously encountered situations and objects might to be the major function of the human sense of smell as opposed to making an essential contribution to finding or recognizing objects. However, in many animals the olfactory sense is one of the primary tools that were developed to make sense of their environment. Thousands of structurally diverse odor molecules perceived and discriminated by vertebrates supply them with a wide range of vital information, ranging from prey and predator localization to mating behavior, underlining the importance of the olfactory sense to the survival of the species.

It comes with no surprise that the fundamental importance of olfaction to life and health was recognized in 2004 by the award of the Nobel Prize in Physiology or Medicine to Drs. Linda Buck and Richard Axel for their pioneering discovery of the olfactory receptors and the understanding of olfactory organization that these groundbreaking findings allowed. Even almost twenty years after the first molecular studies, olfaction is still a very active, popular and productive field of research among the scientific community. The olfactory system

In the olfactory system, sensory information is processed through a series of distinct neuro anatomical structures beginning with the binding of odorant molecules to odorant receptors at the olfactory epithelium in the nose, and ending at the higher cortical areas of the brain, where a perceptual construct is then generated (Reed 1992; Buck 1996; Hildebrand and Shepherd 1997; Mori 1999; Nakamura 2000).

A. The olfactory system

1. The rodent olfactory system

The olfactory system furnishes elaborate molecular and cellular machineries for detection and discrimination of a vast number of chemical compounds in the environment (Axel 1995). Most mammals, including rodents, have two separate olfactory systems that detect and process two functionally distinct classes of chemicals, volatile odorants and

pheromones, through anatomically segregated neural pathways: the main olfactory system and the vomeronasal (accessory olfactory) system (Fig. VI-1) (Buck 2000; Mombaerts 2004). Volatile odorants are detected by a large repertoire of olfactory receptors (ORs) expressed on the cilia and dendritic knob of ciliated olfactory sensory neurons (OSNs) in the olfactory epithelium (OE), that project their axons to the main olfactory bulb (OB). In contrast, pheromones are detected mostly by two families of vomeronasal receptors (VRs) on microvillous sensory neurons in the vomeronasal organ, and their information is transmitted to the accessory olfactory bulb (AOB).

2. The teleost fish olfactory system

However, in teleost fish the situation is radically different. The fish olfactory system consists of one single olfactory organ containing the different classes of OSNs (Fig. VI-1), which project their axons to one single OB. The types of OSNs are distinct with respect to their anatomy, relative position in the OE, and expression. To date, three different types of OSNs were described to mediate odor detection and discrimination in the teleost olfactory epithelium. Besides the more commonly known ciliated and microvillous cells, the crypt receptor neurons are thought to be also involved in olfactory pathways (Hansen and Zeiske 1998), (Hansen and Zielinski 2005) and there is evidence that these cells respond to odorants (Schmachtenberg 2006).

3. Ciliated versus microvillous sensory neurons

The ciliated OSNs have long dendrites, their somata are situated in the deep basal layer of the OE and then express ORs, whereas microvillous OSNs have short dendrites and are located in the superficial layer of the OE, are known to express V2Rs (Cao, Tanguay et al. 2003); (Morita and Finger 1998; Speca, Lin et al. 1999; Hansen, Rolen et al. 2003; Zeiske, Kasumyan et al. 2003; Hansen, Anderson et al. 2004). Moreover, retrograde tracing experiments in several teleost fish species reported that the two types of OSNs project axons to different regions in the OB (Sato, Miyasaka et al. 2005).

The ciliated OSNs project both to the medial and ventral regions of the OB. These regions are activated by both amino acids and bile salts (Friedrich and Korsching 1997; Hansen, Rolen et al. 2003) and project through the LOT and possibly through the medial bundle of the olfactory tract (mMOT). The LOT can discriminate among amino acids (von Rekowski and Zippel 1993) and is involved in feeding behavior (Hamdani, Alexander et al. 2001; Hamdani, Kasumyan et al. 2001), whereas the mMOT is involved in the alarm reaction (Hamdani, Stabell et al. 2000). The lateral region of the OB is innervated by the microvillous OSNs (Morita and Finger 1998; Hamdani, Alexander et al. 2001; Hamdani, Kasumyan et al.

2001; Hamdani el and Doving 2002; Hansen, Rolen et al. 2003) and microvillous neurons are known to discriminate both amino acids and nucleotides (Friedrich and Korsching 1997; Hansen, Rolen et al. 2003) and probably project through the LOT. Taken together, microvillous neurons appear to be involved in feeding behavior whereas ciliated neurons may recognize alarm substance(s) and possibly other pheromones (von Rekowski and Zippel 1993; Hamdani, Alexander et al. 2001; Hamdani, Kasumyan et al. 2001).

Recent experiments making use of transgenic fish labeled with two different molecular cell markers, OMP and TRPC2 for ciliated and microvillous neurons, respectively, confirmed the data from the retrograde tracing experiments described above (Sato, Miyasaka et al. 2005).

4. Crypt sensory neurons

Teleost fish species also display a third type of OSN in their olfactory neuroepithelium: the crypt cell neurons. Recently, cells with crypt-like morphology have also been found in mammalian olfactory epithelium (Hansen 2006). Crypt neurons may possess a duplex sensory apparatus, as they contain both an apical rim of microvilli and several cilia, which protrude from the bottom of a crypt, hence the naming (*cf*. (Hansen and Zeiske 1998)) (Fig. 1). This cell type is distributed mainly in the apical region of the lamellae (Hansen and Zeiske 1998; Hansen and Finger 2000; Hansen, Rolen et al. 2003; Zeiske, Kasumyan et al. 2003; Hansen, Anderson et al. 2004). Although the exact role of crypt ORNs in olfactory pathways is still uncertain, it has been shown in crucian carp (*Carassius carassius*), that their axons project to a central region in the ventral olfactory bulb (Hamdani el and Doving 2006), whose neurons are activated by pheromones (Lastein, Hamdani el et al. 2006). These bulbar neurons project through the lateral bundle of the medial olfactory tract (lMOT), which mediates reproductive behavior (Weltzien, Hoglund et al. 2003). Thus, crypt cells may express olfactory receptors for sex pheromones, which make them interesting targets for olfactory research.

However, no defined molecular marker has become available for crypt cells, so they have been identified either morphologically at the electron-microscopic level (Hansen and Zeiske 1998) or by immunostaining with an antibody directed against bovine S100A1/B (Germana, Montalbano et al. 2004). It is not known which member of the S100 family, if any, was responsible for this cross-reactivity.

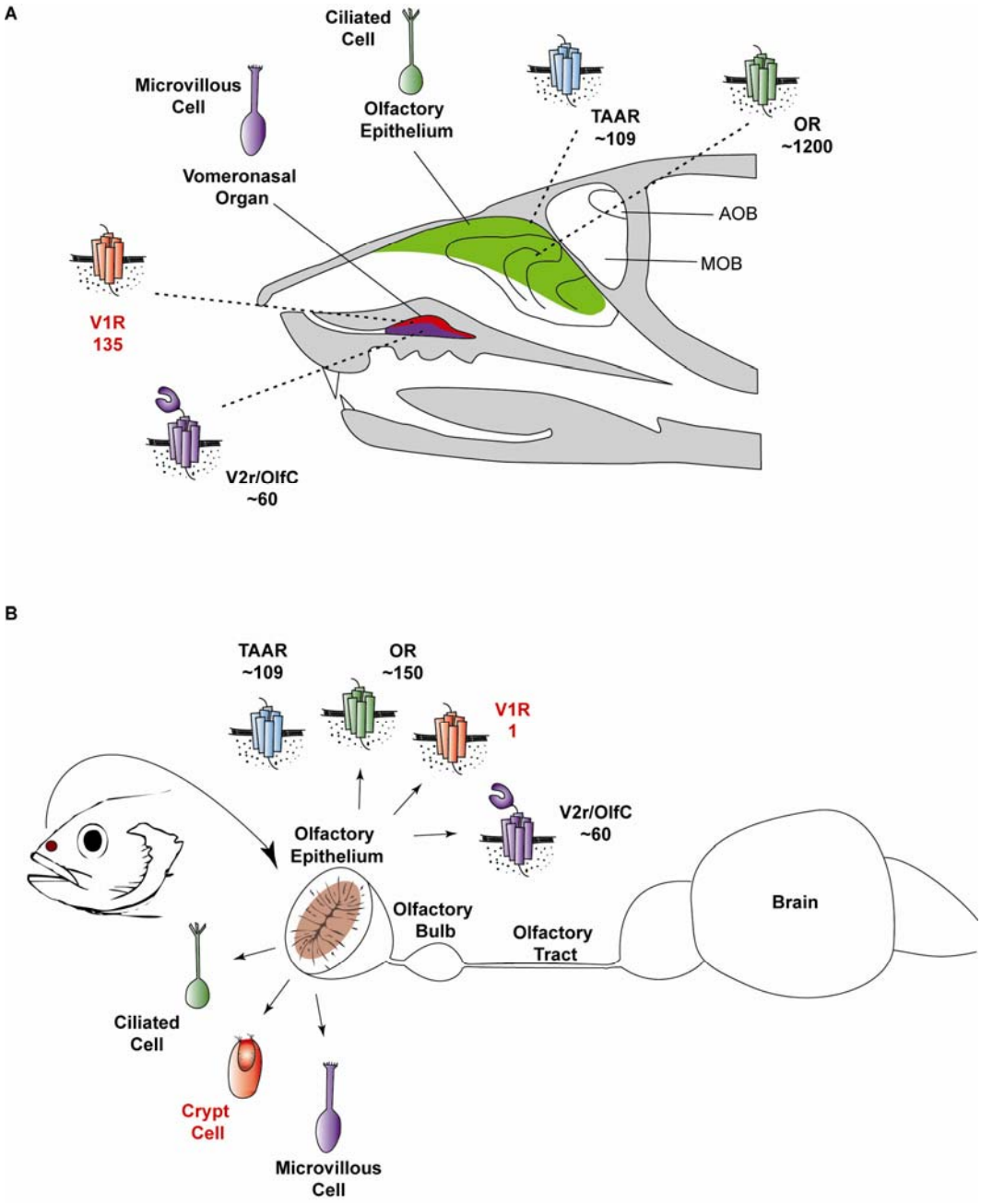


Figure VI-1 | The location of chemosensory organs in the mouse and teleost fish.

(A) A ciliated sensory neuron in the olfactory epithelium of mice expresses either OR or TAAR olfactory receptors. Microvillous neurons in the apical and basal layers of the vomeronasal organ express distinct, unrelated classes of G-protein-coupled pheromone receptors (V1Rs in the apical and V2Rs in the basal layer). (B) A sensory neuron in the olfactory epithelium of teleost fish expresses members of either OR, V1R or V2R olfactory receptors. In fish, in addition to the ciliated and microvillous cells, a third type of neuron is also present – the crypt cells. All the three types of OSNs are located in a single olfactory organ.

B. The olfactory receptor gene family repertoire and its evolutionary dynamics

In 1991, the pioneering work of Linda Buck and Richard Axel unraveled a large and diverse family of G protein-coupled receptors (GPCRs), expressed in the rat olfactory epithelium which were proposed to function as odorant receptors (Buck and Axel 1991). Their work, along with the availability of a large number of fully sequenced genomes, opened the door for several subsequent molecular and bioinformatic studies of the different families of olfactory receptors. A large number of olfactory receptor genes in the same family, as Buck and Axel's originally identified Ors, was described later in other mammals, birds, fish, and amphibians; in contrast to invertebrate species, which display similar, but autonomously expanded families of chemosensory GPCRs. Other independently expanded GPCR families appear to be responsible for pheromone and trace-amine detection in vertebrates. All of these olfactory receptor gene families vary between species. This is not unexpected given that each species considerably have their own characteristic set of chemical signals that are important ought to survival and reproduction. This specificity likely reflects a high level of evolutionary adaptation aimed at creating diversity. The remarkable discriminatory capacity of the chemosensory system (Dryer 2000) is directly linked to the diversity of the olfactory receptor gene families.

1. Olfactory receptors

Humans have about 350 functional odorant receptors (Niimura and Nei 2003), much less than the ~1000-1200 in the mouse and rat genomes, respectively (Gibbs, Weinstock et al. 2004; Young, Shykind et al. 2003). In fish the numbers are several fold smaller, ranging from 42 to 143 putatively functional OR genes in pufferfish and zebrafish, respectively (Alioto and Ngai 2005; Niimura and Nei 2005). Olfactory receptors are indeed one of the largest gene families known, comprising in rat about 6% of their total functional genes, emphasizing the importance of olfaction to the species. It is worth noting that probably a small subset of these genes may have other non-olfactory functions, in addition to or instead of a primary olfactory role. At least one human OR, hOR17-4, is expressed in the testis as well as the nose, responding to the chemical bourgeonal, thus allowing sperm to undergo chemotaxis toward bourgeonal sources (Spehr, Gisselmann et al. 2003). Members of the OR family are expressed mainly in the ciliated neurons (Sato, Miyasaka et al. 2005; Sato, Miyasaka et al. 2007). Members of the OR gene family display monogenic expression, i.e. a particular olfactory sensory neuron expresses only one OR (Serizawa, Miyamichi et al. 2003; Sato, Miyasaka et al. 2007).

The major difference in OR gene family size between species, ranging from 42 to 1430 genes in pufferfish and rat, respectively (Gibbs, Weinstock et al. 2004; Alioto and Ngai 2005; Niimura and Nei 2005) constitutes an attractive starting point to study the evolution of the "olfactome". Comparisons of olfactory gene families of several species revealed that gene birth and death are common in the evolutionary history of these families and had major importance in defining the current total number of genes in these families (Robertson 1998; Young, Friedman et al. 2002; Young and Trask 2002). The most probable cause of gene birth is local gene duplication. Following a duplication event, the resulting copies can follow many evolutionary trajectories. If the new gene is functionally redundant, one of the copies may undergo an inactivating mutation that will remove this gene from the functional repertoire. In contrast, if the new copy accumulates mutations that allow it to recognize a novel, useful odorant molecule, then it is likely that natural selection will favor the retention of the new, modified sequence. A strong selective pressure on creating diversity could even result in positive selection, i.e. higher rates of non-synonymous relative to synonymous substitutions. Conversely, if changes in the sequence eliminate useful ligand-recognition patterns, they would be subject to ''negative or purifying selection'', i.e. the numbers of synonymous substitutions would be more frequent than the non-synonymous ones, as is observed for genes in general.

OR genes occur mostly in clusters, consistent with and early evolutionary origin (Niimura and Nei 2005). Despite this fact, several studies have found evidence for amino acid signatures of positive selection on the olfactory receptors in mammal and fish species (Hughes and Hughes 1993. However, it remains to be seen whether the putatively selected amino acid changes are correlated with a novel gain of function. The dynamic nature of the evolution of this family is characterized by rapid expansion, gene duplication, extensive gene loss via pseudogenization, and diversifying selection (Ngai, Chess et al. 1993; Ngai, Dowling et al. 1993; Gilad, Segre et al. 2000; Waterston, Lander et al. 2002; Young, Friedman et al. 2002; Young and Trask 2002; Alioto and Ngai 2005).

2. Vomeronasal receptors

The mammalian sensory epithelium of the vomeronasal organ is organized in two layers, an apical and a basal layer, each expressing one type of vomeronasal receptor, the V1Rs and the V2Rs respectively (Fig. VI-1) (Buck 2000; Dulac 2000). The sensory neurons of the apical compartment of the VNO express members of the V1R gene family, which transduce signals via a coupled *Gαi* protein; neurons of the basal compartment express members of a second putative pheromone receptor gene family, the V2Rs, which transduce signals via a coupled *Gαo* protein (Dulac and Axel 1995; Herrada and Dulac 1997; Matsunami and Buck

1997; Ryba and Tirindelli 1997; Pantages and Dulac 2000). Like olfactory neurons, each vomeronasal sensory neuron (VSN) expresses only one or at most a few receptors (Roppolo, Vollery et al. 2007), but see (Martini, Silvotti et al. 2001).

3. V2Rs

V2Rs belong to the family C of GPCRs. In rodents the V2R repertoire consists of about 60 genes (Yang, Shi et al. 2005). These receptors are similar to metabotropic glutamate and calcium-sensing receptors in possessing a large N-terminal domain, whereas the C-terminal region contains the 7TM region characteristic for GPCRs (Matsunami and Buck 1997; Ryba and Tirindelli 1997). Based on studies of a goldfish V2R-like molecule, which binds arginine and lysine (Speca, Lin et al. 1999), it appears possible that V2Rs bind proteins rather than volatile organic compounds. This is consistent with observations that the V2R extracellular regions are homologous to bacterial proteins with amino acid-binding properties (Kunishima, Shimada et al. 2000); (Hermans and Challiss 2001). V2Rs are reported to be expressed in microvillous neurons (Sato, Miyasaka et al. 2005).

Similarly to the ORs, analysis of this gene family in several teleost fish species revealed that V2R genes tend to be organized in clusters, though singletons can also be found scattered throughout the genome (Hashiguchi and Nishida 2005). Recent large-scale duplications of V2R-containing chromosomal regions were detected in two V2R gene clusters. The evolutionary dynamics of this family is also characterized by rapid gene turnover and lineage-specific phylogenetic clustering (Hashiguchi and Nishida 2005; Hashiguchi and Nishida 2006). In addition, phylogenetic and comparative genome analyses have shown that the fish and mammalian V2Rs receptors appear to have diverged significantly from a common ancestral gene(s), with these receptors likely mediating chemosensation of different classes of chemical structures by their respective organisms (Alioto and Ngai 2006; Hashiguchi and Nishida 2006).

4. V1R

The apically expressed V1R family has 135 members in the mouse genome organized into several gene clusters, with genes sharing from 25% to nearly 100% amino-acid sequence identity (Rodriguez, Del Punta et al. 2002). Their phylogenetically closest relatives are the T2R (taste receptor type 2) receptors (with 15-20% amino-acid sequence identity), suggesting that the genes encoding these two receptor families derived from a common ancestral gene. Similar to the OR genes, V1Rs display a 1 kilobase, intronless genomic structure (Buck and Axel 1991). In 2002, Boschat and colleagues (Boschat, Pelofi et al. 2002) identified 2-heptanone, a putative pheromone, as a ligand for one member of the V1R

family (V1Rb2). In another study, V1R genes could be linked to reproductive behavior: Transgenic mice, in which a 600 kb gene cluster containing 16 different V1R genes was deleted, showed a reduction of maternal aggressive behavior. Furthermore, the VNO of these knockout mice did not respond to three out of eight putative tested pheromones. Curiously, even though the deleted gene cluster includes V1Rb2, the vomeronasal organs of the knockout mice nevertheless respond to 2-heptanone, suggesting that multiple receptors might recognize and respond to this compound (Del Punta, Rothman et al. 2000). Members of the V1R family are expressed in the microvillous neurons and co-express Gαi1/2 in mammals (Dulac 2000; Pantages and Dulac 2000). While several studies classified the V2R and OR receptors as evolutionary old families, with about 50 to 150 members already present in several fish species (Hashiguchi and Nishida 2005; Niimura and Nei 2005), the V1R receptor family was considered a recent family that originated from a single V1R-like receptor gene in fish (Pfister and Rodriguez 2005) or alternatively from a single gene pair (Shi and Zhang 2007). Although species-specific expansion and loss of genes and even whole subfamilies is a recurrent theme in all three mammalian receptor families (Lane, Young et al. 2004; Zhang, Rodriguez et al. 2004; Grus, Shi et al. 2005), as well as in fish ORs (Niimura and Nei 2005) and fish V2R related gene families (Hashiguchi and Nishida 2005; Alioto and Ngai 2006), the V1R expansion from a single gene pair to over a hundred genes in some mammalian species appears somewhat extreme. I therefore undertook to examine the V1R-like gene repertoire in fish.

5. TAAR receptors

Trace amine-associated receptors (TAARs) are close relatives of G protein-coupled aminergic neurotransmitter receptors (Lindemann and Hoener 2005). Initially, TAARs were considered neurotransmitter receptors based on the expression and effects of some family members in the central nervous system (Lindemann and Hoener 2005). However, Liberles and Buck reported that several mammalian *taar* genes (some of which they could deorphanize) are expressed in olfactory sensory neurons (Liberles and Buck 2006). Thus, the *taar* genes joined a growing number of GPCR families that serve as olfactory receptors (cf. (Buck 2000). Following the cloning of the first TAAR receptors in mammals (Borowsky, Adham et al. 2001), TAAR genes have been found in lower vertebrate genomes (Gloriam, Bjarnadottir et al. 2005) and recently it has been suggested that the family occurs already in lamprey (Hashiguchi and Nishida 2007). The first *in silico* study uncovered about 50 TAAR receptors in zebrafish and 8 in a pufferfish (Gloriam, Bjarnadottir et al. 2005), but these numbers were nearly doubled in a follow-up study (Hashiguchi and Nishida 2007). Nevertheless, the delineation from classical aminergic neurotransmitter receptors has not

been investigated thoroughly, and consequently both the scope of the TAAR family and its evolutionary origin have not become clear. TAAR genes were shown to co-express G*α*Olf, suggesting that they are expressed at least in ciliated neurons (liberles and buck). In this thesis I have analyzed both the scope and the evolutionary history of the TAAR gene family in fish.

C. Olfactory signaling transduction

Olfactory sensory neurons transduce odor signals by coupling GPCRs to one or more downstream effector molecules. As mentioned above, GPCRs activate heteromeric GTP binding proteins and intracellular second messengers (Axel 1995) (Firestein 1992; Firestein 1996; Firestein, Breer et al. 1996; Zhang and Firestein 2002) (Elsaesser and Paysan 2005; Elsaesser and Paysan 2007) (Schild and Restrepo 1998). Two main excitatory transduction mechanisms are described to be used in OSNs, one mediated by cyclic AMP, and the other by IP3; although the last one is to date subject of controversy (Gold 1999) and references therein) and may in fact be a modulating influence, not an independent pathway (Elsaesser and Paysan 2007). Cyclic nucleotide signaling is best understood in vertebrate olfactory receptor neurons (Fig. VI-2). In that pathway odorants activate ORs (buck 2000) on the cilia of olfactory receptor neurons (ORNs) and, by way of a Gα protein (Gαolf), stimulate an adenylyl cyclase (type III) to synthesize adenosine 3,5-cyclic monophosphate (cAMP) (Schild and Restrepo 1998) (Matthews and Reisert 2003; Reisert, Bauer et al. 2003). cAMP opens a cyclicnucleotide–gated (CNG) cation channel to produce a membrane depolarization (Zufall, Firestein et al. 1994; Schild and Restrepo 1998; Matthews and Reisert 2003; Reisert, Bauer et al. 2003). Influx of Ca^{2+} through the CNG channel opens a Ca^{2+}-activated chloride (Cl) channel, leading to Cl^{-} efflux and further depolarization (Schild and Restrepo 1998) (Matthews and Reisert 2003; Reisert, Bauer et al. 2003). Simultaneously, the Ca2þ influx decreases cAMP synthesis and the effective affinity of CNG channels for cAMP, both effects producing olfactory adaptation (Schild and Restrepo 1998) (Matthews and Reisert 2003; Reisert, Bauer et al. 2003).

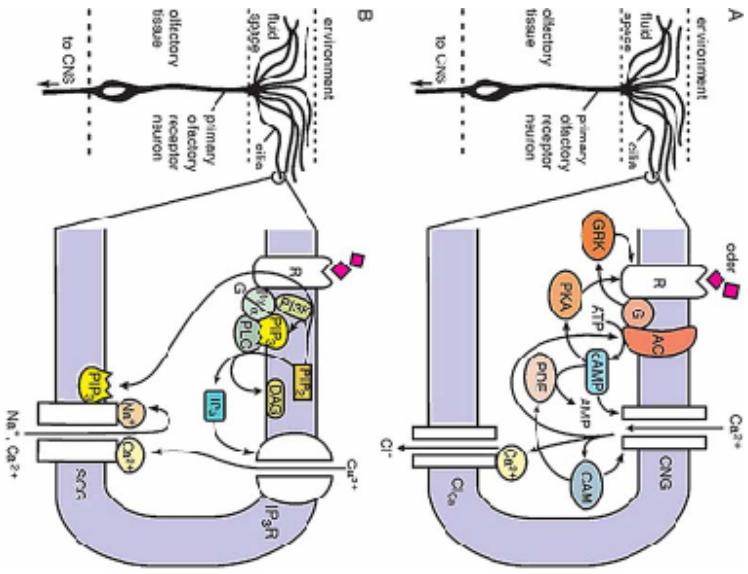


Figure VI-2 | Two Intracellular Signaling Cascades Implicated in Chemosensory Transduction

(A) Diagram of cyclic nucleotide signaling in the transduction compartment (olfactory cilia) of vertebrate olfactory receptor neurons. Odorant molecules bind to a receptor protein (R) coupled to an olfactory specific Gs-protein (G) and activate a type III adenylyl cyclase (AC), increasing intracellular cAMP levels. cAMP targets an olfactory-specific cyclic-nucleotide gated ion channel (CNG), a nonselective cation channel that increases intracellular calcium and secondarily activates a calcium-activated chloride channel thought to carry the majority of the transduction current. Other, regulatory pathways are also shown. (B) Diagram of phosphoinositide signaling as currently understood in the transduction compartment (outer dendrite) of lobster olfactory receptor neurons. Odorant molecules bind to a receptor protein (R) coupled to a Gq-protein and activate both phospholipase-C (PLC) and phosphoinositide 3-OH kinase (PI3K) to generate diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP3), and phosphatidylinositol 3,4,5-trisphosphate (PIP3), respectively, from phosphatidylinositol 4,5-bisphosphate (PIP2). We assume, therefore, that PIP3 in concert with ‘‘release’’ of extracellular calcium from a plasma membrane IP3 receptor (IP3R), also known to be expressed in the transduction compartment, target a lobster homolog of a transient receptor potential channel, a nonselective cation channel that is modulated by both sodium and calcium (SGC) and that has been shown to carry the majority of the transduction current. Details of these pathways vary in other receptor cells and other species. Adapted from (Ache and Young 2005).

According to the inositol triphosphate (IP_3) model a different subset of odorants activates a different subset of receptor proteins that couple via a Gαq/11 or Gαi/o heterotrimeric G protein to phospholipase C (PLC). PLC cleaves the headgroup of the membrane phospholipid phosphotidyl inositol, producing diacylglycerol (which remains in the membrane) and IP3 (which is water soluble). The IP3 model suggests the existence of IP3-gated Ca2+ channels in the ciliary membrane that mediate Ca^{2+}-influx and membrane depolarization (Restrepo, Teeter et al. 1996; Rebecchi and Pentyala 2000). Moreover, PLC stimulation mediated by Gαi and Gαo, seems to be effected via the Gbetagama complex of these heterotrimeric G-proteins (Rebecchi and Pentyala 2000), Gbetagama subunits associated with Gαi and Gαo in VSNs may play an important role in transduction processes. A recent study by Runnenburger and colleagues identified the Gβ subtypes as well as the

Gδ subunits expressed in the VNO, notably in the two groups of chemosensory neurons and the authors hypothesize that PLC activation in the two populations of chemosensory VNO neurons is mediated by different Gβδ complexes (Runnenburger, Breer et al. 2002). Similarly to mammals, fish also present evidence for the existence of both AMP and IP3 transduction pathways (Miyamoto, Restrepo et al. 1992; Miyamoto, Restrepo et al. 1992; Restrepo, Boekhoff et al. 1993; Restrepo, Okada et al. 1993; Restrepo, Okada et al. 1993). Later, G*α*−protein subunits were described to be present in catfish OSNs (Abogadie, Bruch et al. 1995; Abogadie, Bruch et al. 1995) (Dellacorte, Restrepo et al. 1996). Hansen and coworkers (2003) showed in catfish that cyclic AMP mediates olfactory sensory transduction in response to amino acids and bile acids. The G*α*olf/s protein was localized to the cilia and plasma membrane of ciliated OSNs of the channel catfish, while the microvillous and crypt neurons seem to use G*αq/11 and* G*α*o, respectively (Hansen, Rolen et al. 2003), goldfish (Hansen, Anderson et al. 2004), and in the perciform round goby (Belanger, Smith et al. 2003). Moreover, evidence for an IP3 mediated pathway in response to amino acids has been shown in salmon (Lo, Bradley et al. 1993) and zebrafish (Ma and Michel 1998). In zebrafish, the ciliated cells were shown to express the cyclic nucleotide gated A2 subunit together with members of the OR family. Microvillous cells were found to express the transient receptor potential channel C2 together with members of the V2R-like family. Moreover, both types of OSNs project to distinct areas of the olfactory bulb (Sato, Miyasaka et al. 2005). The use of genetic, physiological and morphological probes offers promising avenues of progress towards the elucidation of the mechanisms of signaling transduction in the teleost OSNs, which still remain poorly understood.

D. Aims:

Indeed, the sense of smell is so important that exerts sufficient selective pressure to maintain several hundreds of olfactory receptor genes throughout evolution. Studying the evolutionary dynamics of the olfactory receptor genes is thus an instrumental approach to better understand the olfactory sense. This approach requires both the identification and characterization of the different olfactory receptor gene families not only in mammals but also in lower vertebrates.

For two of the four mammalian olfactory receptor gene families, very incomplete data had been available. I report here the identification and characterization of fish V1r-like and *taar* gene families and continue with an analysis of cellular expression, co-localization and signal transduction cascade components.

VII. RESULTS

The nature of olfactory stimuli has dramatically changed during the water-to-land transition of vertebrate evolution. Corresponding evolution of olfactory receptor gene families is to be expected. While mammalian olfactory receptor gene families are well characterized, not much is known about the total repertoires and evolutionary dynamics of some of these families in teleost fish.

I have uncovered and characterized a new class of olfactory receptor genes in fish and contribute to the delineation and characterization of another olfactory receptor gene family in fish. With this thesis I present both the classification of these gene families including some cell-biological aspects and compare their properties to those of their mammalian counterparts, thus elucidating the transition in the olfactory receptor gene repertoire upon the teleost-tetrapod transition and consequent colonization of the terrestrial environment.

A. Ora family

1. Four novel V1R-like genes were identified in each of five teleost species

A recursive search strategy starting with all known *V1r* genes (see Methods for details) and using automatic ortholog annotation in combination with the tBlastN algorithm (for details see Materials and Methods) uncovered four novel genes in the zebrafish genome (Fig. VII-1). All orthologs of all four *Danio rerio* genes could be identified in four further fish species, *Gasterosteus aculeatus, Oryzias latipes, Takifugu rubripes,* and *Tetraodon nigroviridis* (three-spined stickleback, medaka and two pufferfish, respectively). No pseudogenes were detected in any of the five teleost species.

I proposed to name these *V1r*-like genes *ora* (olfactory receptors related to class A, cf. Schioth and Fredriksson 2005; www.gpcr.org/) and not *V1r*, since that name refers to the tetrapod vomeronasal organ which fish do not possess - all the fish olfactory receptors are expressed in the main olfactory epithelium. The designation '*ora*' reflects both the olfactory-specific nature of these receptors as well as their phylogenetic position within the GPCR super family. Individual *ora* genes were numbered from 1 to 6, beginning with the first zebrafish *ora* gene to be identified (Pfister and Rodriguez 2005). The new nomenclature has been accepted by the zebrafish nomenclature committee and is shown in Fig.32. It mirrors the recent re-naming of the fish *V2r*-like genes as *OlfC* genes (cf. Alioto and Ngai 2006). In the phylogenetic comparison with other teleost chemosensory receptor families (see below) *ora* genes emerge as monophyletic group.

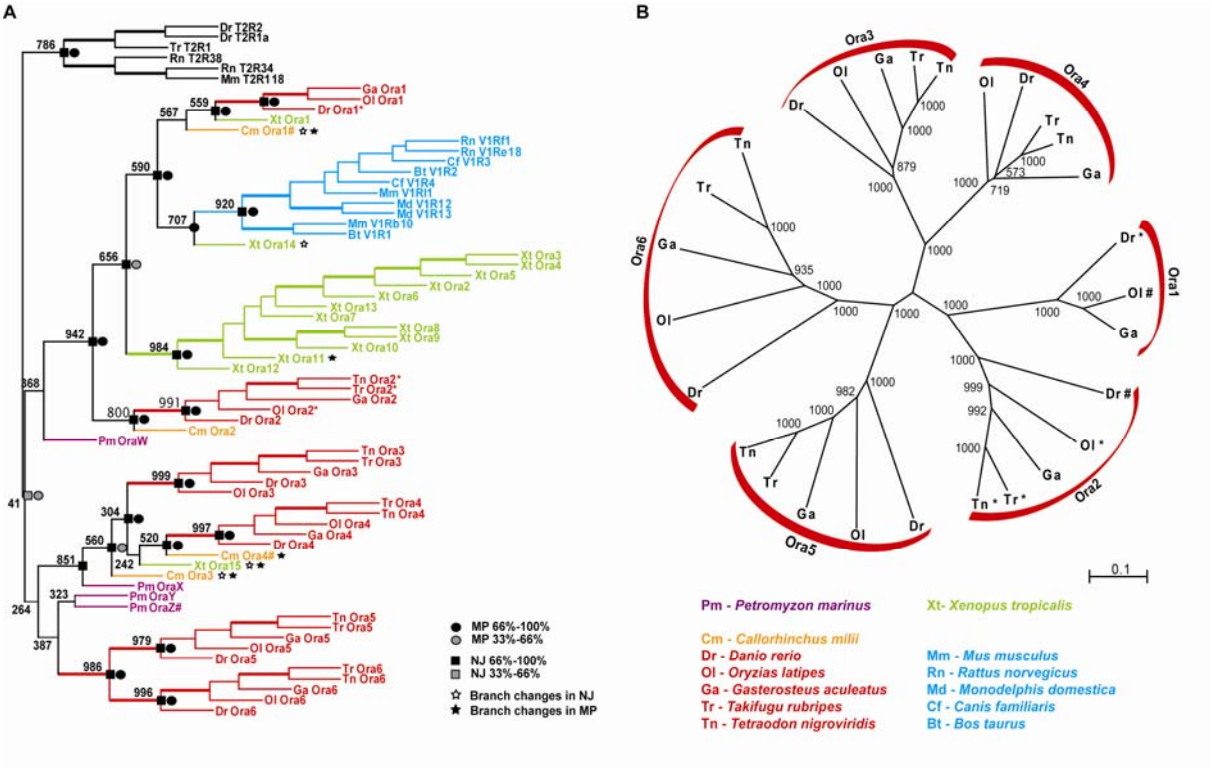


Figure VII-1 | Phylogenetic tree of the fish Ora family

(A) 28 fish Ora (red), 4 shark Ora (orange), 4 lamprey Ora (violet), 15 frog Ora (green) with some mammalian V1R (light blue) representatives and T2R (orange) as the closest relatives. Tree presented constructed using the ML method. (B) 28 fish *ora* genes. Tree presented constructed using the NJ method. Bootstrap support (total 1000 replications) is indicated at the major nodes. Scale bar indicates the number of amino acid substitutions per site. Asterisk, see (Pfister and Rodriguez 2005); double cross, see (Shi and Zhang 2007). *ora*: olfactory receptors of class A; T2R: putative taste receptors of type 2 (Ishimaru et al. 2005); V1R: vomeronasal type 1 receptors (Grus et al. 2005). The V1R receptors are a subset of V1Rs from all mammalian organisms annotated in the NCBI database (mouse, rat, human) and described in publications (opossum, cow, dog, (Grus et al. 2005). The phylogenetic position of the full mammalian V1R set of annotated and published genes is identical (cf. Supplementary Figure 1). Bold lines indicate ML bootstrap values greater or equal to 85%. Unfilled and filled stars indicate branch changes in the NJ and MP methods, respectively. Black and grey circles and squares indicate that the clades downstream of the node are supported by bootstrap values of 66-100% and 33-66% for the trees constructed using MP and NJ methods, respectively.

2. The ora genes form a single clade together with mammalian V1Rs

Using the T2Rs as outgroup; I compared the newly found genes to their most closely related chemosensory receptor families, the mammalian *V1r* genes (Fig. VII-1A). With respect to T2R, OR (Fig. XI-1) and other chemosensory receptor gene families (T1R, OlfC, data not shown) all fish *ora* genes form a monophyletic clade, supporting their identification as a single family separate from the other chemosensory receptor families. The Ora clade includes all mammalian V1R receptors (Fig. VII-1A, Fig. XI-1); thus the Ora family can be considered paraphyletic, with the mammalian V1Rs originating as a single subclade within the Ora family. Both mammalian and fish taste receptors of the T2R family (Ishimaru, Okada et al. 2005) are close phylogenetic neighbors, but nevertheless clearly segregate from the

Ora family with high bootstrap values (Fig. VII-1A). Odorant receptor genes (ORs) are even more distant from the *ora* genes (Fig. XI-1).

The six teleost *ora* genes subdivide into three pairs, *ora*1-*ora*2, *ora*3-*ora*4, and *ora*5-*ora*6. In the phylogenetic tree these subclades are supported by maximal bootstrap values (Fig. VII-1B). Orthologs of the individual genes in all cases can be identified unambiguously, again with maximal bootstrap values (Fig. VII-1A). Conserved amino acids are mostly restricted to the orthologs of a single gene, but often a particular position is conserved in all orthologs of a gene pair (cf. Fig. VII-2), consistent with the branchpoint pattern of the phylogenetic tree. Motifs conserved across two or all three gene pairs are comparatively rare, as detailed below.

3. Ora family size in jawless and cartilaginous fish close to that of teleost fish

I also analyzed the available databases for the presence of *ora* orthologs in one species of cartilaginous fish – *Callorhinchus milii* – and in one jawless fish – *Petromyzon marinus*. Four novel Ora family members were identified in each of the above mentioned species (Fig. VII-1).

This repertoire for these two species is probably not yet complete due to the following reasons: in the case of *Callorhinchus milii*, only 1.4 fold genomic coverage is currently available and in the case of *Petromyzon marinus* the genome assembly is defective due to both a repetitive genome rich in A/T and a high heterozygosity rate. Thus the total number of *ora* genes in cartilaginous and jawless fish (four genes present in each species) may possibly be rather similar to that of teleost fish. Since in the case of the sea lamprey genes no clear direct homologs were delineated, a new rule of nomenclature was adopted: these genes were named provisionally with W, X, Y and Z; not to foreclose a coherent nomenclature for *ora* genes of basal taxa, which will only be possible after completion of the respective genome projects.

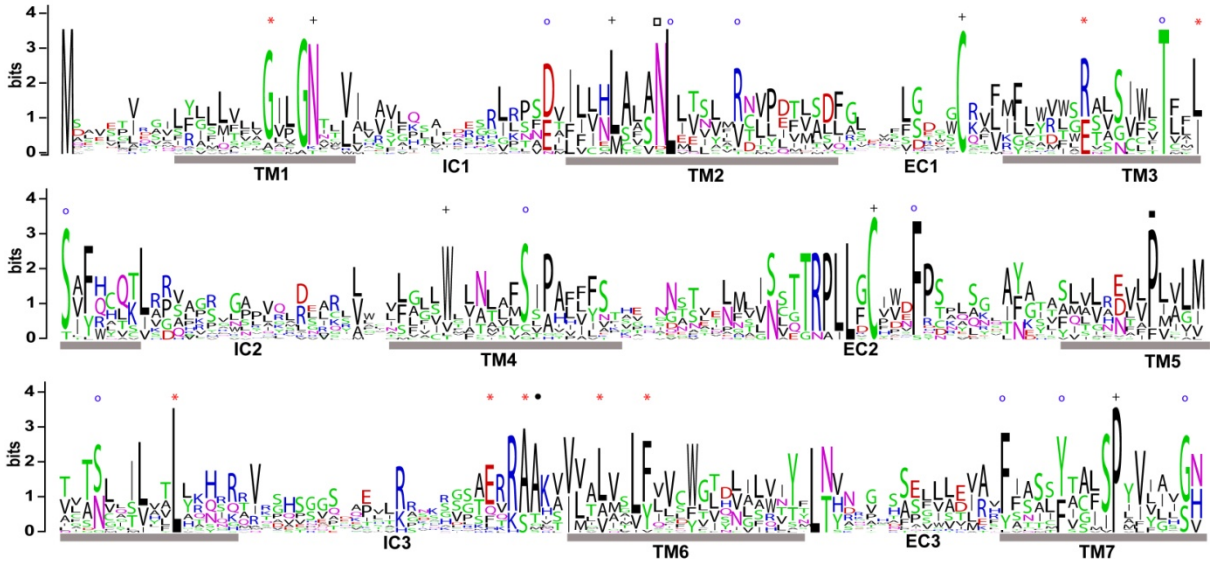


Figure VII-2 | Conserved sequence motifs of the Ora family.

Conservation of predicted amino acid sequence for the fish Ora repertoire is displayed as a sequence logo. In this representation, the relative frequency with which an amino acid appears at a given position is reflected by the height of its one-letter amino acid code in the logo, with the total height at a given position proportional to the level of sequence conservation. The regions corresponding to the transmembrane (TM) domains and the extracellular and intracellular domains (EC & IC) are numbered and indicated. Sequence alignments were manually edited (for details see Methods section). Of fourteen motifs conserved in V1Rs (all of them single amino acids, identified by (Rodriguez et al. 2002) eight are not conserved in ORs (cf. Niimura and Nei, 2005) and consequently were chosen as analytical criterion here (asterisks). Crosses represent residues generally conserved among other GPCR families and circles represent residues conserved in fish *ora* genes, but not in mammalian V1R genes. Filled circles represent residues conserved between Ora and TAARs. Squares represent residues conserved between Ora, V1R and TAARs. Filled squares represent residues conserved between Ora, T2R and TAARs.

4. Low overall similarity but high degree of conservation for motifs characteristic for mammalian V1Rs

The *ora* genes constitute a highly heterogeneous family, with homologies often as low as 15% and minimally 11% (Fig. XI-3). I therefore analyzed the retention of characteristic sequence motifs in order to obtain a second line of evidence supporting the identification of *ora* genes both as a new family, and as a V1R-related family.

Mammalian V1Rs are already quite divergent and comprise e.g. in mice twelve distantly related subfamilies (Rodriguez, Del Punta et al. 2002). Consequently there are very few defining sequence motifs, all of them single amino acids, which are retained in nearly all family members. Initially, fourteen such motifs were reported to be V1R-specific (Rodriguez, Del Punta et al. 2002). However, six of them are present also in several other general class A GPCR-specific, e.g. OR, TAAR and T2R thus cannot be used to delineate the V1R family (Fig. VII-2). Furthermore, these six motifs are located either inside the TM (one “N” in TM1, one “L” in TM2, one “W” in TM4 and one “P” in TM7) or EC (one “C” in both EC1 and EC2) domains and at least some of them are known to play a instrumental role in the general

structural stability of the molecule (Rodriguez, Del Punta et al. 2002). The remaining eight are V1R-specific and are also highly conserved in the fish *ora* genes (Fig. VII-2), supporting the assignment as V1R-like genes. The degree of conservation is highest for the *ora*3-*ora*4 gene pair, slightly lower for the *ora*1-*ora*2 gene pair, and lowest (though clearly significant) for the *ora*5-*ora*6 gene pair, consistent with its larger distance from the mammalian V1Rs in the phylogenetic tree.

Beyond these V1R-specific motifs plus the above described general class A GPCR-specific motifs, *ora* genes contain ten Ora-specific conserved amino acids located mainly in the TM domains. All these Ora-specific residues are - with three exceptions;"L" in TM2, "P" in TM5 and "A" in IC3 (Fig. VII-2) - not conserved in either fish OR, TAAR or T2R genes (see Ishimaru, Okada et al. 2005; Niimura and Nei 2005), and thus distinguish the Ora family from the odorant, trace-amine and the taste receptor families (Fig. VII-2). In teleost fish, three of these motifs ("T" and "S" in TM3 and "F" in EC2) are conserved in all *ora* genes, two ("S" in TM4 and "F" in TM7) in 5 of 6 genes, and five ("D" in IC1, "R" in TM2, "S/N" in TM5 and "Y" and "G" in TM7) in 2 of 3 *ora* gene pairs (the differing gene pair is variantly *ora*1-*ora*2, *ora*3-*ora*4, or *ora*5-*ora*6).

This degree of motif conservation among the teleost fish *ora* genes extends to the *ora* genes of both cartilaginous and jawless fish species analyzed. The shark shares all but one of the described general GPCR and Ora-specific motifs with the teleost fish species as well as the lamprey. Instead of an "S" in TM4, Cm ora4 displays an "A". Lamprey *ora* genes display slightly more variability in two of the specific Ora motifs, "S" in TM4 and "F" in EC2 (Fig. VII-2).

In mammalian V1Rs some of these motifs are weakly conserved, and others are not conserved at all, supporting the assignment of the fish *ora* genes as a family separate from, but related to mammalian V1Rs.

5. ora genes precede teleost speciation with an early origin in jawless fish

The presence of Ora orthologs in the five distantly related teleost fish species suggests by itself an ancient evolutionary origin of this family, before the segregation of all five teleost species, at least 350 million years ago (Nakatani, Takeda et al. 2007). The discovery of orthologs in cartilaginous and jawless fish sets the origin of the family close to the origin of the vertebrates.

In teleost species, Ora orthologs are without exception closer related than paralogs (Figs. VII-1,3), suggesting that all six Ora family members are evolutionary older than the speciation events in the teleost lineage taken into account here. Moreover, for each *ora* gene the phylogenetic tree closely reflects the phylogenetic relationships of the species. The two

pufferfish studied belong to the same family (*Tetraodontidae*) and their orthologs are related closer to each other than to those from any other species. The stickleback and medaka orthologs constitute their next neighbors, as pufferfish, stickleback and medaka belong to three subdivisions of the same superorder *Acanthopterygii* (orders *Tetraodontiformes*, *Gasterosteiformes*, and *Beloniformes*, respectively). In all cases the zebrafish orthologs occupy the most distant position (zebrafish being the evolutionarily most distant fish in this comparison, as it is an *Ostariophysi*, not a *Neoteleostei* like the four other species). For zebrafish *ora*1 orthologs were found in stickleback and medaka, but not in the two pufferfish species. Since stickleback, medaka and pufferfish belong to the same superorder *Acanthopterygii*, but zebrafish doesn't, this appears to be a case of gene loss in the pufferfish family, possibly related to the extreme reduction in genome size so characteristic of that family.

The clear separation in three teleostean subclades (Fig. VII-1) suggests the presence of three ancestral genes in the most recent common ancestor (MRCA) of teleost and tetrapods. As already mentioned, both the jawless (sea lamprey) and cartilaginous (elephant shark) fish species analyzed in this study have an *ora* repertoire consisting of four genes each. The elephant shark *ora* genes are clear orthologs of the *ora1, ora2, ora3* and *ora4* genes present in the teleost species analyzed, thus the equivalent *ora* gene pairs must have been formed before the segregation of cartilaginous from bony fish. The sea lamprey *ora* genes have no direct orthologs neither in cartilaginous nor teleost fish (Fig. VII-1) and occupy a very basal position in the three *ora* subclades, thus placing the origin of the Ora family in the MRCA of jawless and jawed fish; firstly with the Pm *oraW* being the ancestral gene for the *ora*1-*ora*2 clade, secondly with the Pm *oraX* as the ancestral gene for the *ora*3-*ora*4 clade and lastly with the remaining two genes Pm *oraY* and Pm *oraZ* as ancestors of the *ora*5-*ora*6 clade (Fig. VII-3). Since both Pm *oraY* or Pm *oraZ* are direct paralogs, it is not clear whether one of this genes was lost in the transition from jawless to jawed fish or, alternatively, one was gained by a local gene duplication of the other after the origin of the *Gnathostomata* lineage.

6. Gene loss and gene gain upon transition to tetrapods

Mammalian V1R genes all belong to the *ora*1-*ora*2 clade, and all mammalian genes form a separate subtree within that clade (Figs. VI-1, VII-1A, XI-1 and data not shown). These results are consistent with a loss of two clades (*ora*3-*ora*4, and *ora*5-*ora*6) somewhere after the teleost/tetrapod split, as well as a massive expansion of the remaining clade (Fig. VII-3). Such course of events is very reminiscent of the evolution within the OR gene

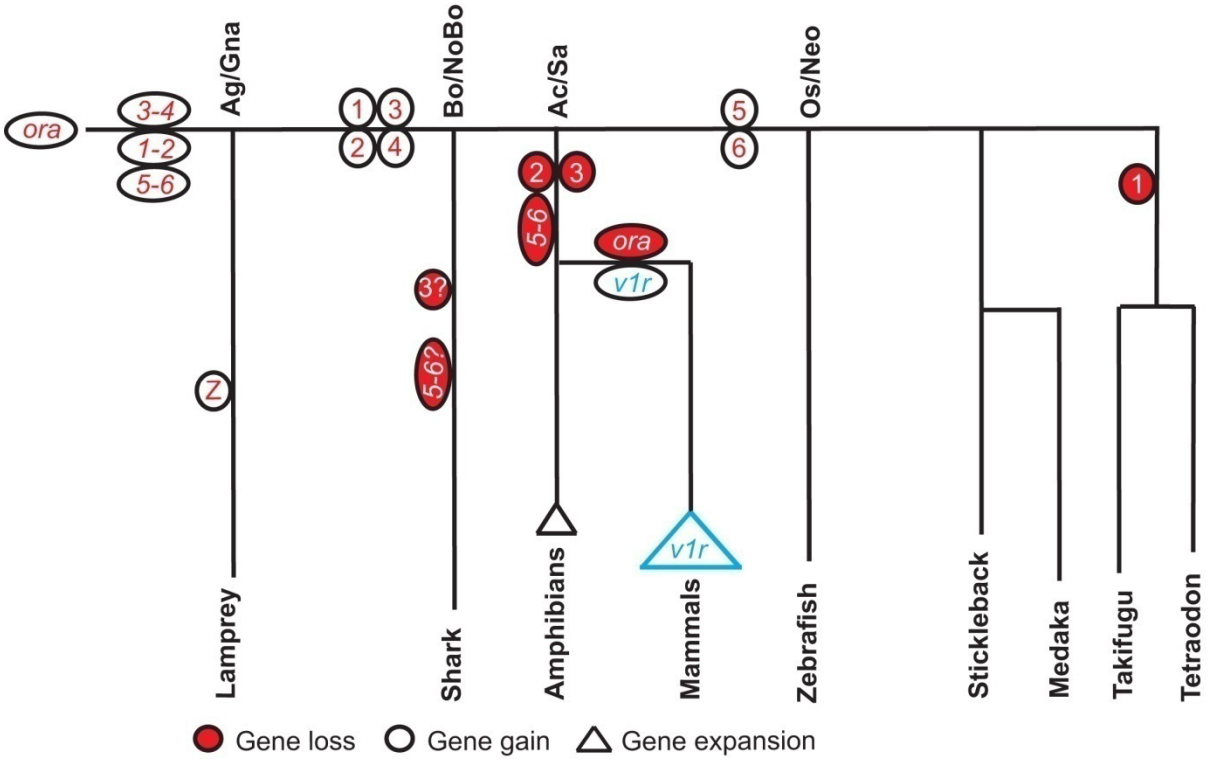


Figure VII-3 | Estimated minimal evolutionary age of Ora clades and genes.

Open circles represent the gene gain events in each lineage, and red-filled circles represent the gene loss events. Inside each circle is the name of the respective gene(s) or clade. Emergence of the ora gene family, of the three clades of ora genes and of the v1r gene family is indicated by ovals. Red-filled ovals represent the loss of a whole clade or family. Triangles represent gene expansions. The major phylogenetic transitions are indicated: Ag/Gna, jawless/jawed fish segregation; bo/nobo, bony fish/cartilaginous fish; ac/sa, actinopterygian/sarcopterygian split; os/neo, ostariophysii/neoteleostei segregation. The maximum parsimony principle was followed, thus gene gains are depicted at the last possible stage before additional gains would become necessary for explanation, but may in fact have occurred earlier.

family, where nearly all subfamilies present in teleosts disappeared in mammals, and exactly one subfamily underwent a massive expansion (Niimura and Nei 2005). To clarify the time scale of the Ora evolution I have searched the genome of an amphibian tetrapod for *ora* genes. In *Xenopus tropicalis* I found 15 *ora* genes in total, which belong to two of the three fish clades (Figs. VII-1, XI-1 XI-2). Clade *ora*5-*ora*6 is missing, and clade *ora*3-*ora*4 is only represented by a single gene, Xt *ora*15. However, clade *ora*1-*ora*2 contains a single ortholog of *ora*1, another isolated gene (Xt *ora*14) closest to the nodal point of origin of this clade, as well as a large expansion of highly related genes (Xt *ora*2-13), well within the range of such expansions in mammalian species (Fig. VII-3). With respect to gene loss Xenopus represents an intermediate stage, but with respect to gene gain it resembles the mammalian situation. Due to their phylogenetic position it appears appropriate to name the Xenopus genes as *ora*, since they intermingle with fish *ora*, but not with mammalian *V1r* (Figs. VII-1 and VII-3). Numbering begins with Xt *ora*1, because the ortholog assignment is

unambiguous only in this case, following numbers are given according to the position in the NJ phylogenetic tree (Saraiva and Korsching 2007).

I hypothesize that the partial gene loss might be due to the loss of a fully aquatic life style in amphibians, and that the gene expansion is related to the transition to a terrestrial environment, consistent with a major shift in function of *ora* genes during this transition.

7. Strong negative selection for ora genes, but no evidence for positive selection

To better understand both the high degree of intraspecies variability and the high degree of interspecies conservation of *ora* genes I analyzed the evolutionary constraints that are acting on this gene family.

Paralog homology is usually below 25% amino acid identity and often as low as 15%, with average values for each species close to 20% (Fig. VII-4B). Most of this divergence is due to radical amino acid changes, since even omitting conservative amino acids exchanges, the average similarity of paralogs is still below 40% for all species (Fig. VII-4C). Homology within paralog gene pairs is somewhat higher than between them, with *ora*5 vs. *ora*6 comparisons always resulting in lower values than those obtained for the other two gene pairs, ora1 *vs.* ora2 and ora3 *vs.* ora4.

Ortholog homologies are much higher and in fact identity of any ortholog pair is higher than that of any paralog pair in all possible pairwise comparisons bar one (Fig. XI-2). The average identity in all ortholog comparisons is 60%, with mean values for the individual *ora* genes ranging between 46 and 72%, and average ortholog similarities go up from 60% to 81%, with an average value for all *ora* genes of 72% (Fig. VII-4).

These values appear large enough, cf. (Wolfe and Sharp 1993) to allow analysis of non-synonymous (dN) vs synonymous (dS) substitutions as a means to calculate the selective pressures acting on the *ora* genes. When the number of dN equals the number of dS, the dN/dS ratio equals 1, which corresponds to neutral selection. If the number of non-synonymous changes is higher than the number of synonymous changes, then dN/dS>1,which indicates positive selection. On the other hand, if the number of synonymous (dS) changes is higher than the number of non-synonymous changes, then dN/dS<1 and we are in the presence of negative selection (Nei and Gojobori 1986).

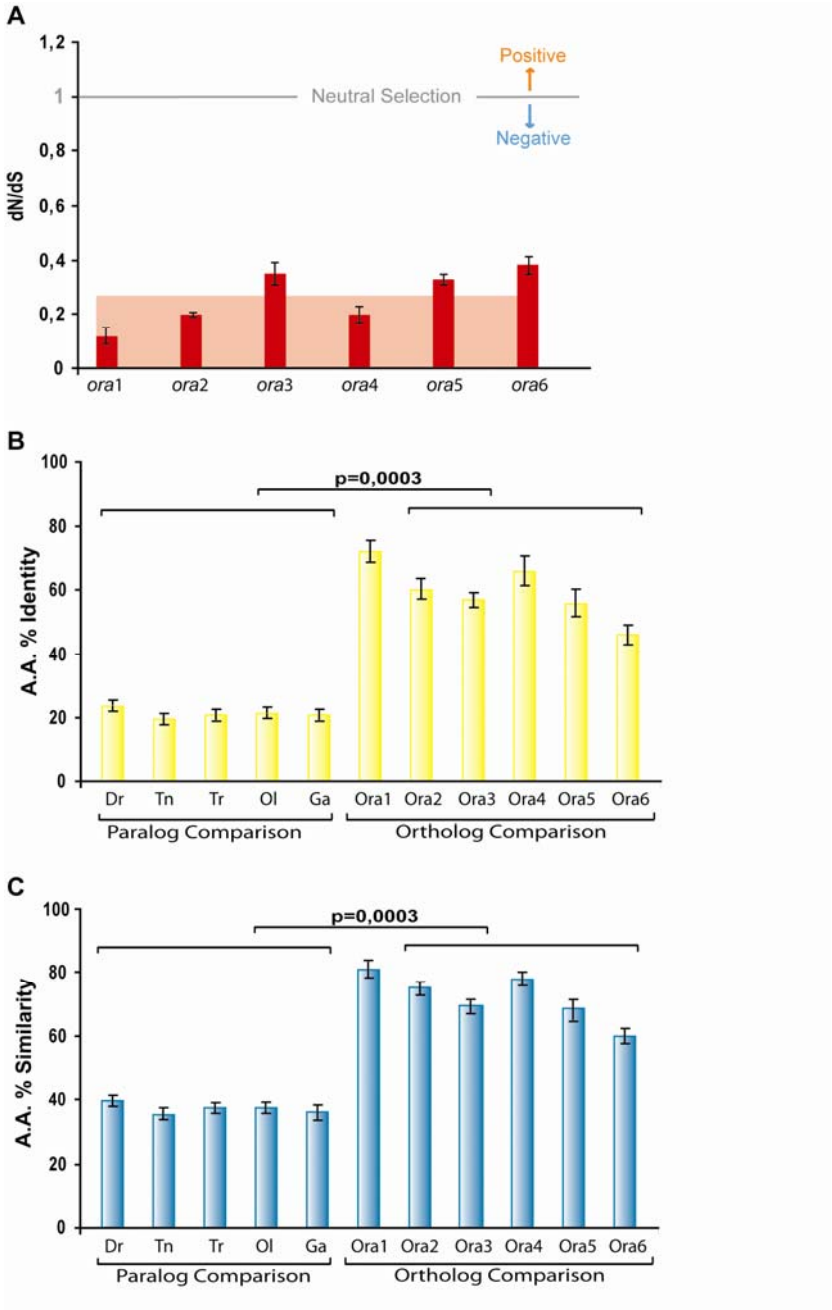


Figure VII-4 | Identity, similarity, and conservation level of the ora genes.

(A) dN/dS ratios of the six *ora* genes. For each gene, the dN/dS ratio was determined for all possible pairwise comparisons between orthologs. The average value is indicated by background shading. and the mean value was plotted. (B) Amino acid % identity is calculated for each gene comparisons between orthologs) or species (paralog comparisons) by averaging the values of all possible pairwise comparisons inside the described group. (C) Amino acid % similarity is calculated for each gene or species by averaging the values of all possible pairwise comparisons inside the described group. A, B, C). The bars correspond to the associated standard deviation as measure of the variance within each group. B, C) Differences between values for ortholog and paralog comparisons are highly significant.

To avoid distortion of the dN/dS ratio by beginning saturation of synonymous substitutions (Gojobori 1983) the dS values should not exceed a certain value, differently given as 2 or 3 (Mank, Axelsson et al. 2007). I therefore analysed the dS values for each of the pairwise comparisons separately and verified that 49% of the dS values (26 values) are below 1.0, 32% (17 values) are between 1.0 and 1.5, 15% (8 values) are between 1.5 and 2.0 and only 2 values are slightly above 2.0 – the highest being 2.2 (Data not shown). Since nearly no dS values are above 2 and the vast majority is even below 1.5, we assume that saturation of the synonymous substitutions does not distort the overall dN/dS calculations. I observed a very low average dN/dS ratio for comparisons between orthologs (0.25), with values for individual genes ranging between 0.11 for *ora*1 and 0.37 for *ora*6 (Fig. VII-4A). All values clearly indicate strong negative selection, i.e. the *ora* genes are slowly evolving genes. A low dN/dS value of the *ora* genes together with a high divergence between *ora* genes indicates a very ancient origin of this slowly evolving gene family, consistent with the phylogenetic analysis' results. This is drastically different from the properties of the mammalian V1R family, which is characterized by fast evolution and consequently highly species-specific gene repertoires. Incidentally, this difference in evolution rates may be related to the difference in pseudogene frequency: high numbers of pseudogenes are present in the V1R family (Zhang, Rodriguez et al. 2004; Grus, Shi et al. 2005), but none were detected in the Ora family. Since overall strong negative selection could mask positive selection at a few individual codon sites, I also analysed the dN/dS ratio for each sequence position, using a manually optimized alignment of all orthologs for a particular *ora* gene. This analysis was performed separately for all the ortholog *ora* groups. As expected from the summary dN/dS analysis, extended regions of the coding sequence show evidence for moderate to strong negative selection (Fig. VII-5). Moreover, no evidence for a positively selected site was found in any of these genes. A comparison between the ortholog groups shows rough similarity in the pattern of negative selection, although no specific motifs could be identified between genes (Fig. VII-5).

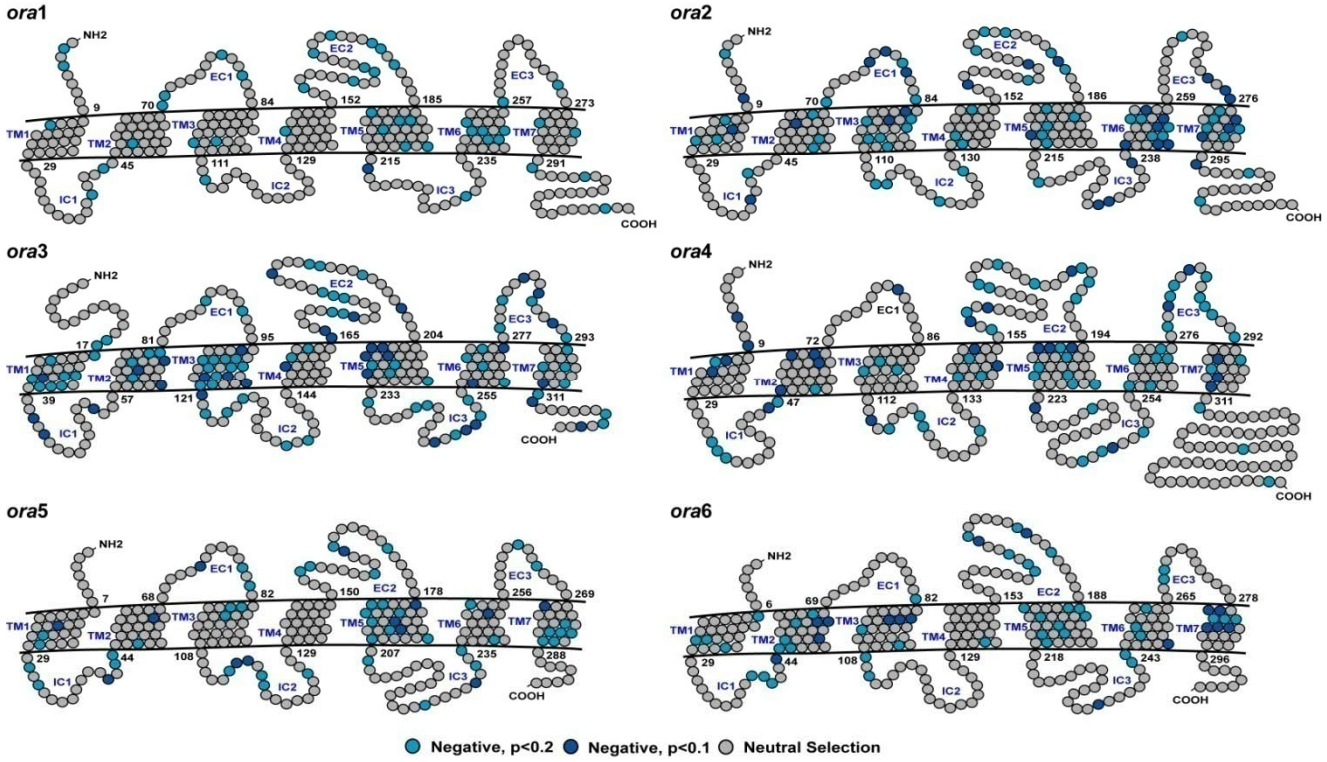


Figure VII-5 | Sites under positive and negative selection in *ora* coding sequences.

A schematic representation of site-by-site selective pressure is shown on *ora* receptor sequences drawn based on the edited nucleotide alignment, which was generated from the corresponding amino acid alignment used for Figure 4. SLAC analysis shows the probability of sites being under selective pressure (negative selection in light-blue ($p<0.2$) or blue ($p<0.1$), neutral selection in gray, positive selection not observed even at $p<0.2$ level. All orthologs of each gene were used for this analysis; the results all the teleost genes are shown.

8. Mono and multiexonic gene structures are present in the ora family

It is generally believed that all mammalian *V1r* genes possess a single exon structure of about 1 Kb in size (Dulac and Axel 1995; Saito, Mimmack et al. 1998; Rodriguez, Feinstein et al. 1999; Del Punta, Rothman et al. 2000; Grus, Shi et al. 2005). In concordance with this, a monoexonic structure of about 1000 bp in size is shared by all the *ora* genes that belong to the clades ora1-ora2 and ora5-ora6 (Fig. VII-6). This monoexonic structure is also shared by all of the intact three sea lamprey (Pm oraW, Pm oraX and Pm oraY) and two elephant shark *ora* genes (Cm ora2 and Cm ora3), which all were retrieved from the correspondent genomic DNA databases. Also the Pm oraZ, although incomplete, is most likely intronless, as it lacks only the TM7 and the C terminal domains. The other incomplete *ora* gene is Cm ora1, which has both teleost and lamprey intronless genes as orthologs i.e. is observed already in the jawless fish ancestral *ora* gene of the Ora1-Ora2 clade and therefore is expected to share the intronless gene structure.

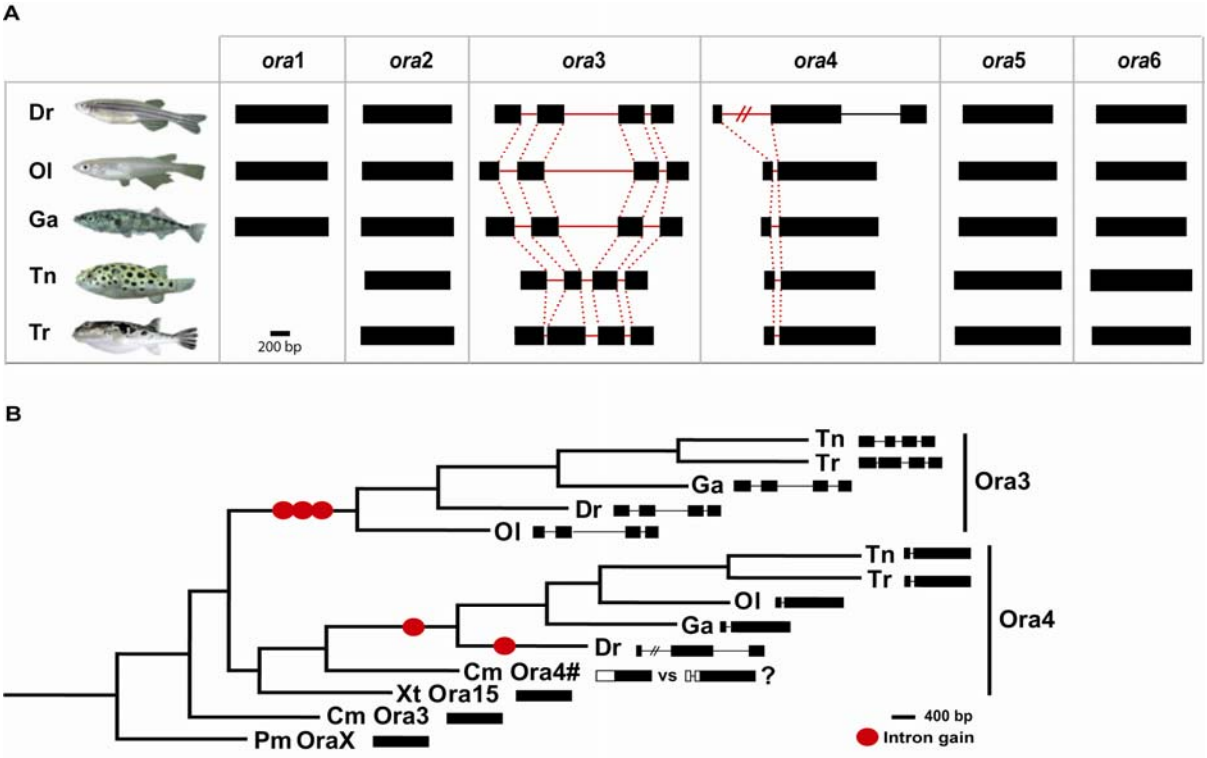


Figure VII-6 | Genomic structure and intron dynamics of the 28 fish *ora* genes.

(A) Predicted exon/intron structure is drawn to scale for all the *ora* genes: six zebrafish *ora*s (Dr; *Danio rerio*), six medaka (Ol; *Oryzias latipes*), six stickleback (Ga; *Gasterosteus aculeatus*), five fugu (Tr; *Takifugu rubripes*) and five tetraodon (Tn; *Tetraodon nigroviridis*) *ora gene*s. (B) Using maximum parsimony, predictions for all independent events of intron gain or loss are depicted in the phylogenetic tree detail. Exons are represented by the black filled rectangles. Conserved and non-conserved introns are represented by the red and black line connecting the exons, respectively. Red filled circles indicate an intron gain event.

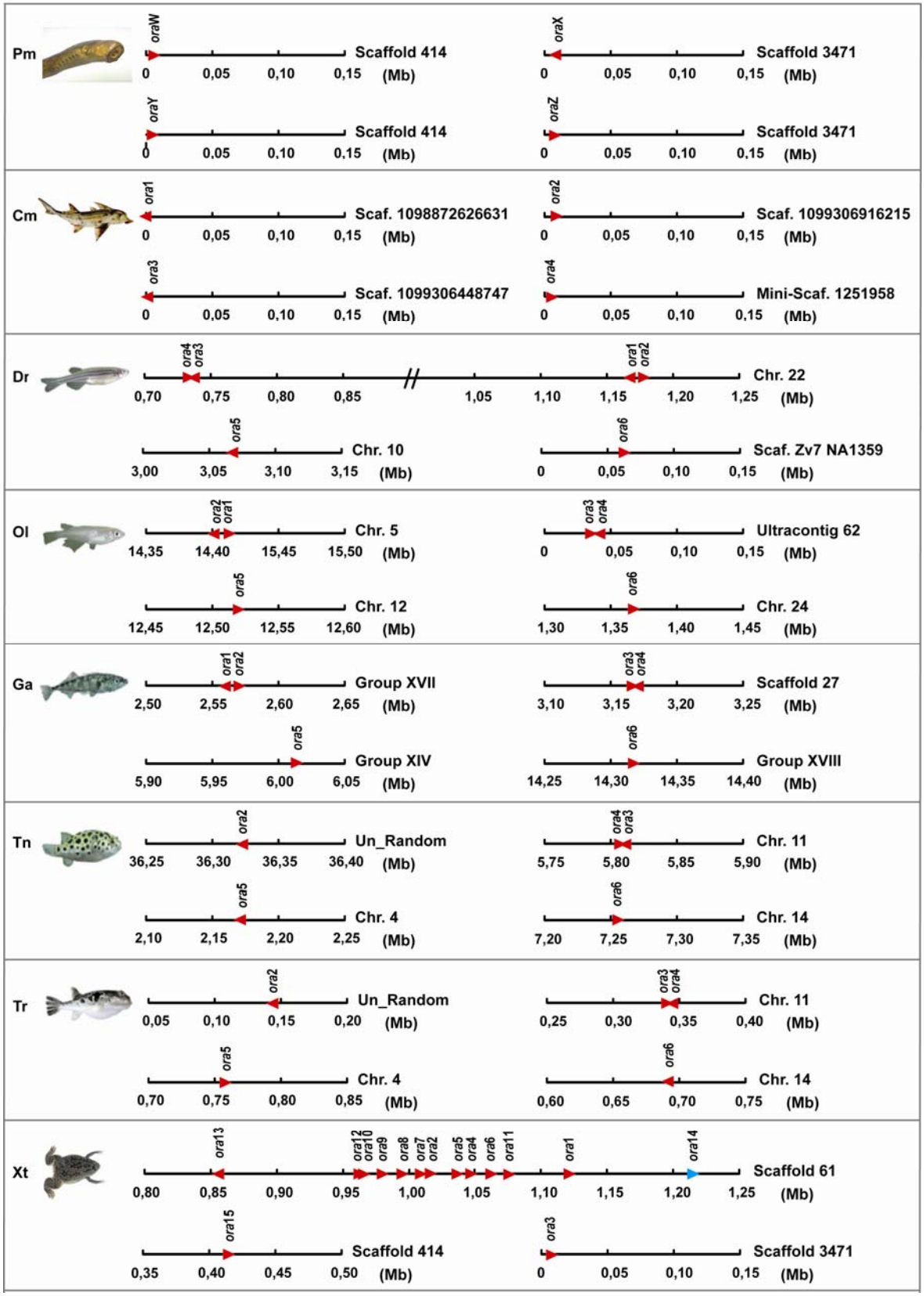
It came then as a surprise that about one third of the fish *ora* genes show a multi-exon structure in the coding region with two, three, or four exons (Fig. VII-6). These multiexonic genes all belong to the Ora3-Ora4 gene pair, with the *ora*3 exhibiting a four-exon gene structure and the *ora*4 showing a two-exon gene structure conserved across all teleost species. Additionally, the Dr *ora*4 exhibits one more intron in the C terminal (Fig. VII-6). The gene structure of the elephant shark ora4 gene still remains to be elucidated, since it lacks the N terminal-TM3 part of the molecule where the intron/exon border is found to be present in the teleost fish orthologs.

9. Evolutionary history of intron gains in the teleost ora gene family

The prevailing *ora* genomic structure is, as described above, intronless. This structure is maintained in the mammalian relatives of the Ora1- Ora2 clade (intronless genes in fish), given that all the tetrapod V1r-like genes reported to date also appear to be monoexonic (Dulac and Axel 1995; Grus, Shi et al. 2005). This appears to be the ancestral state, since the lamprey oraW and oraY genes are intronless as well (Figs.. VII-1 AND VII-6).

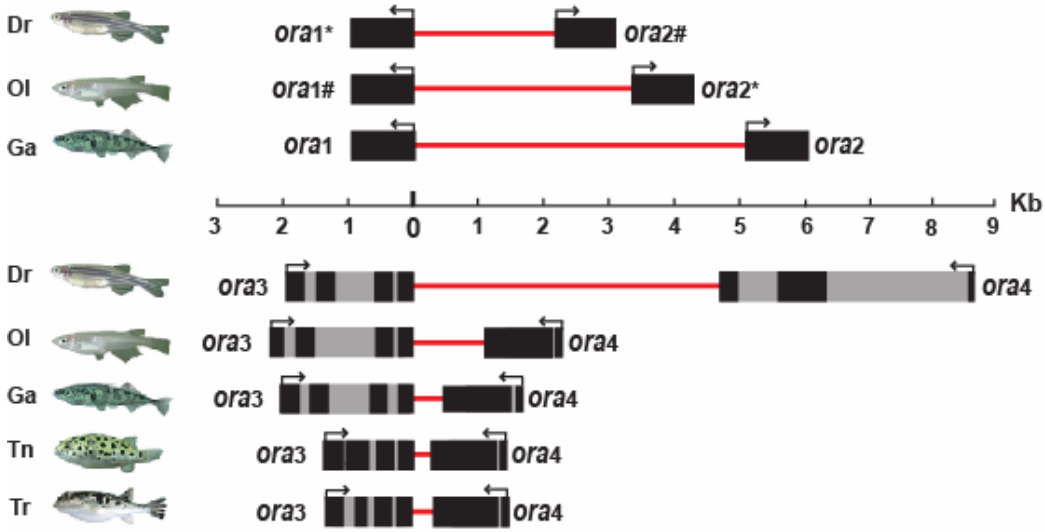
In marked contrast, *ora*4 possesses two exons in four fish species (three in *Danio rerio*), and in all cases a small N-terminal exon is followed by a several fold larger C-terminal exon. For *ora*3 four approximately equal-sized exons are predicted in all five fish species examined (Fig. VII-6). For the zebrafish representatives these exon predictions have been confirmed by sequencing all products of the RT-PCR analysis (Fig. VII-9A). Furthermore, *ora*3 intron/exon borders are exactly conserved between all five teleost species and the same holds true for *ora*4 (Fig. VII-6). The sole intron/exon border in *ora*4 does not correspond to any intron/exon border in *ora*3 (Fig. VII-6A). The striking conservation of particular intron/exon borders suggests that putative conserved non-coding elements might be involved in regulation of *ora* gene expression. The presence of three conserved introns in ora3, is best explained by independent events early in evolution of teleosts, after segregation from cartilaginous fish, but before teleost speciation (Fig. VII-6B). Similarly, the presence one conserved introns in ora4, might also be explained by independent events after segregation from cartilaginous fish and before teleost speciation. However, since the shark Cm ora4 sequence is incomplete in its N terminal, I cannot exclude the possibility that the intron gain could also have occurred already in cartilaginous fish. Furthermore, the three intron gains in ora3 must have happened independently of that in ora4, i.e. after the local gene duplication which led to this gene pair consistent with the phylogeny results (see above). The most parsimonious explanation for the presence of an additional intron in the zebrafish ora4 is a second gain that might have occurred later in the *ostariophysi* fish subsequent to the segregation from the *neoteleostei* lineage.

Figure VII-7 | Genomic arrangement of the ora1-ora2 and ora3-ora4 gene pairs.



10. Four ora genes are arranged in closely linked gene pairs in head-to-head and tail-to-tail orientation

Between the three clades Ora1-Ora2, Ora3-Ora4, and Ora5-Ora6 there is no genomic linkage apparent. This holds true also for the gene pair *ora*5-*ora*6, as these two genes occur on different chromosomes or at least in different contigs in all species studied. However, the gene pairs *ora*1-*ora*2 and *ora*3-*ora*4 exhibit a striking pairwise arrangement with very short intergenic distance of a few kilobases, in the case of the pufferfish less than 1 kb (Fig. VII-8). In zebrafish these two gene pairs are located in a single cluster of about 0.5Mb in chromosome 22 (Fig. VII-7). The orientation is head-to-head (head means 5' end) for the *ora*1-*ora*2 gene pair, and tail-to-tail (tail means 3' end) for the *ora*3-*ora*4 gene pair. This pairwise arrangement and even the orientation is conserved without exception in all five teleost species examined (Figs. VII-7 and VII-8). Thus, a significance of this arrangement for regulation of expression may be assumed. This would result in an evolutionary constraint and consequently slower evolution than for the non-linked *ora*5, *ora*6 genes. Indeed, ortholog identities are always minimal for *ora*6, and nearly always minimal for *ora*5



compared to the remaining four genes, consistent with an accelerated evolution of both *ora*5 and *ora*6.

Figure VII-8 | Genomic locations of all the ora genes.

ora genes can be found both in clusters as in singletons on different chromosomes. Horizontal scale, distance in Mb. Orientation of the genes is indicated by left/right pointing position of the triangle (reverse/forward strand, respectively). The frog ora gene marked in blue is the one that clusters basally to the mammalian V1Rs.

11. Loss of linked gene pair in frog concurrent with rapid expansion in the ora1-ora2 clade

Physical maps of the *ora* genes in all the fish species analyzed are shown in Fig. VII-7. The exact position of 17 genes out of the total 51 genes could be determined in the zebrafish, medaka, fugu and tetraodon genomes. The remaining genes are assigned either to scaffolds (sea lamprey, elephant shark, zebrafish, stickleback and frog), groups (stickleback) or contigs (medaka) (Fig. VII-7) due to the incomplete status of the respective genome projects.

In sea lamprey and elephant shark, nothing can be concluded about their possible distribution in clusters and/or singleton since the exact physical position of the genes cannot be determined due to the initial stage of their genome projects. In frog, despite the fact that the physical maps of its genome are also not yet available, it is clearly visible and noteworthy that thirteen out the total 15 *ora* genes are sitting in a cluster of about 0.4 MB in scaffold 61 (Fig. VII-7). All this genes belong to the ora1-ora2 clade (Figs. VII-1 and VII-7) and are consistent with the previously mentioned gene expansion upon the transition to a terrestrial environment. In frog, this gene expansion was probably caused by a series of local gene duplications in the Ora1-ora2 clade. The other two clades show no expansion and were lost either partially (ora3-ora4) or completely (ora5-ora6) (Fig. VII-3). Consistent with an origin by local gene duplications the phylogenetic position of the frog *ora* genes mostly parallels their respective genomic locations (Figs. VII-1, VII-7 and XI-1).

12. Specific expression of all ora genes in the adult olfactory epithelium

Any olfactory receptor is expected to be expressed in olfactory receptor neurons situated in the olfactory epithelium. To test that prediction I performed RT-PCR with seven different tissues and all zebrafish *ora* genes. All six genes were expressed specifically in the

olfactory epithelium (Fig. VII-9A), supporting their assignment as olfactory receptors. None of the genes was expressed in the taste cell-containing tissues barbels and lips (Fig. VII-9A), confirming the segregation of the *ora* gene family from its closest phylogenetic neighbors, the T2R family of taste receptor genes. For a higher spatial resolution of the expression patterns at the cellular level, in situ hybridization performed for all six *ora* genes. All *ora* genes are expressed in sparse cells mainly located in the apical layer of the lamellae and confined to the sensory region of the zebrafish olfactory neuroepithelium (Figs. VII-9B, C); supporting the assumption that *ora* genes are olfactory receptors.

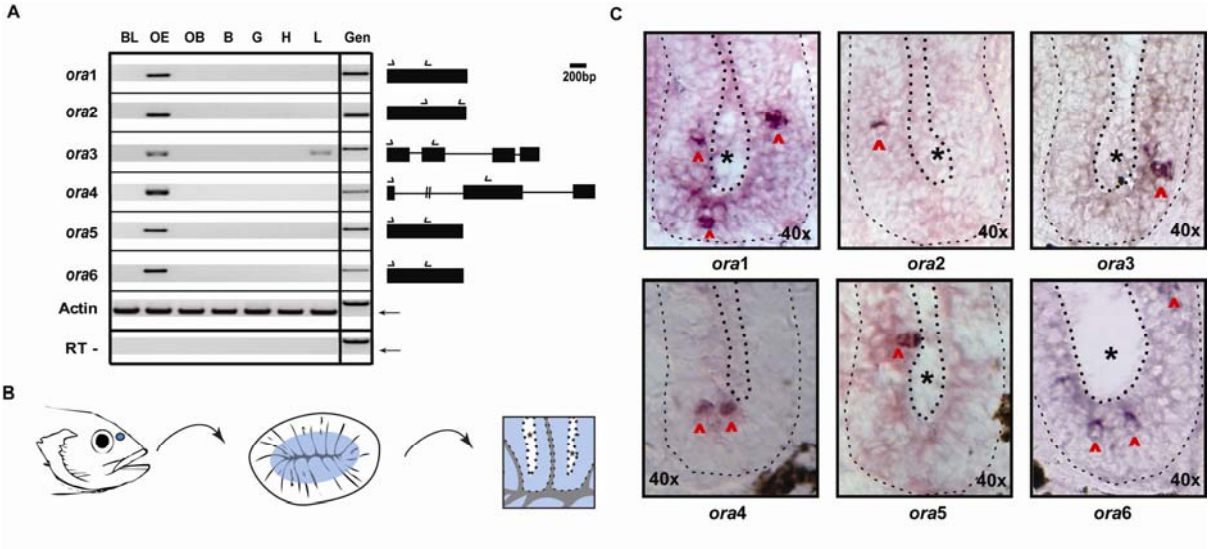


Figure VII-9 | Expression of ora transcripts in the zebrafish olfactory system.

(A) Expression of *ora* mRNA detected by RT-PCR. PCR amplifications were performed by using gene specific primers (arrows above the gene structure scheme). BL, barbels+lips; OE, olfactory epithelium; OB, olfactory bulb; B, brain; G, gills; H, heart; L, liver, and Gen, genomic DNA. Actin, both plus and minus RT, and genomic DNA as template for all *ora*s and actin were used as controls. The single actin band as well as the absence of actin amplification products in the 'minus RT' condition confirm the absence of genomic DNA contamination. That genomic DNA, if present, would have generated a visible amplification product, is shown in the lane labeled 'Gen'. Gel sections shown all correspond to the 400-900 bp range, with exception of the much larger genomic product of *ora*4. The weak band with *ora*3 in liver cDNA might be due to minor ectopic expression, as has been reported for several olfactory receptor genes (e.g. (Vanderhaeghen, Schurmans et al. 1993). Arrows in the actin rows point to the expected position for the cDNA product. (B) Schematic representation of the localization of the OE followed by a drawing of a horizontal section of OE (lamellae are cut perpendicular to their flat face) and finally an enlargement of two lamellae. The central blue-colored area in the lamellae indicates the location of the sensory neuroepithelium (cf. Weth et al., 1996), grey lines, basal lamina, grey jagged spots, lumen. (C) In situ hybridizations with *ora*1, *ora*2, *ora*3, *ora*4, *ora*5 and *ora*6 in horizontal sections of the OE, with antisense RNA probes. The black asterisks indicate the lumen. Each half-lamella is enclosed by dashed lines, thicker in the apical region and thinner in the basal region adjoining the basal lamina. Red arrowheads point to the labeled neurons.

13. In the embryo, ora genes show broader expression including but not limited to the olfactory placode

To assess the developmental onset of the *ora* genes, whole mount *in situ* hybridization of zebrafish larvae 5 days post fertilization were performed. At this stage zebrafish have completed organogenesis and major behavioral patterns are already functional. Three *ora* genes – ora1, ora3 and ora5, representing the three *ora* clades – were analyzed and showed a broader spatial distribution than one might have expected from the RT-PCR results (Fig. VII-9 and VII-10). Nevertheless expression in the olfactory placode was detected in the three cases, thereby setting the onset of the *ora* genes at least on the fifth day post fertilization (Figs. VII-10D, H and M). beyond the olfactory placode all three *ora* genes were found to be expressed in the branchial arches (Figs. 12C, F and K), thymus for ora1 and

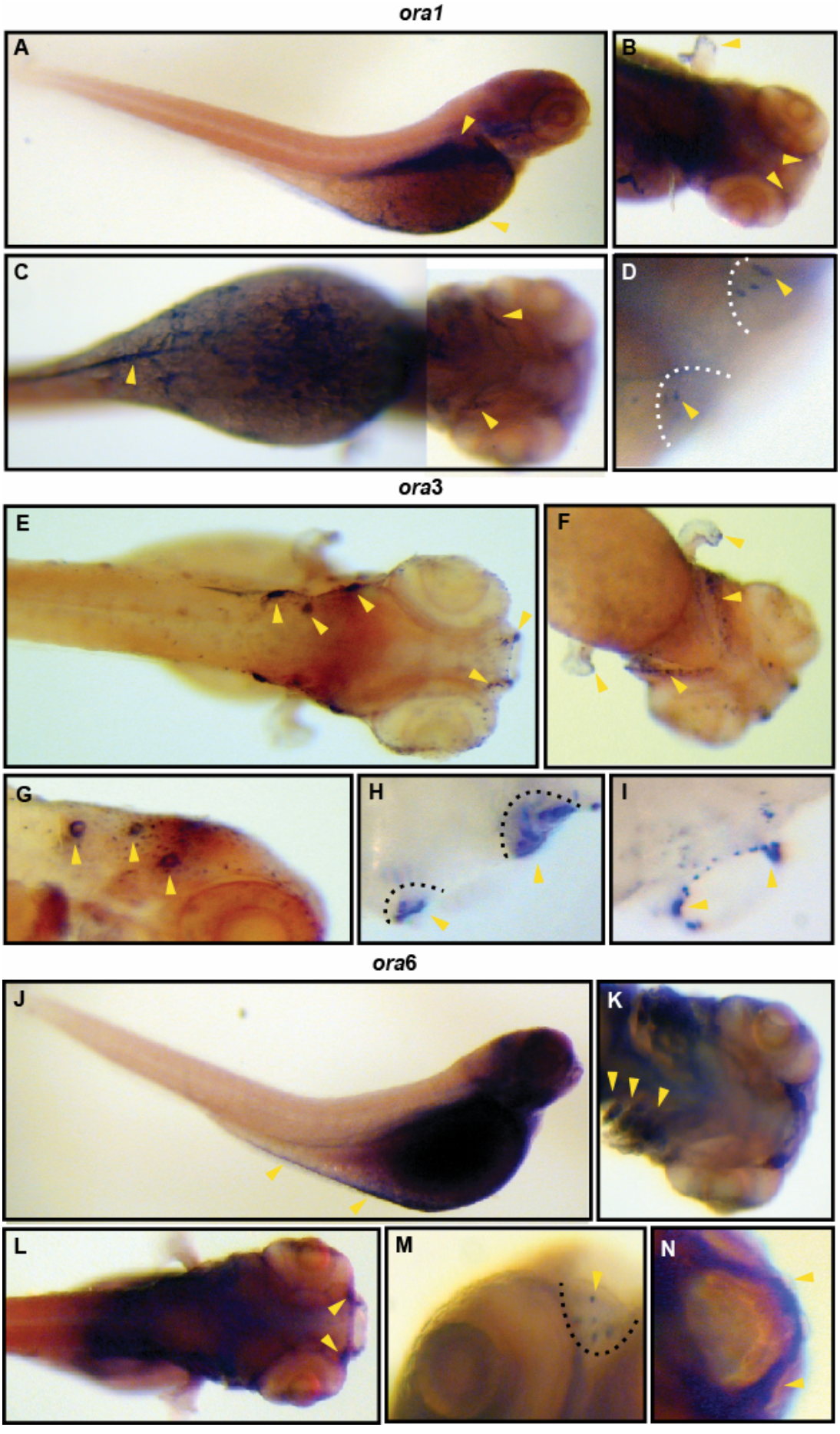

ora1
A
B
C
D
ora3
E
F
G
H
I
ora6
J
K
L
M
N

Figure VII-10 | *ora* genes expression pattern by whole mount in situ hybridization in 5dpf old zebrafish larvae.

Five day old zebrafish larvae were hybridized with RNA antisense probes for A-D) ora1, E-I) ora3 and J-N) ora6. (A) Lateral view shows expression of ora1 in the thymus; (B) Dorsal view, olfactory epithelia and tip of pectoral fins (arrows) express ora1; (C) Ventral view (anterior to the right) shows expression in the yolk region and in the branchial arches; (D) Enlarged view of the olfactory epithelia (delimited by white dashed lines) showing labeled individual neurons by ora1 (arrows). (E) Dorsal view showing expression of ora3 in the olfactory epithelia, mouth region and cells surrounding neuromasts in the head region (arrows); (F) Ventral view showing expression of ora3 in one pair of branchial arches and tips of pectoral fins (arrows); G) Enlarged view of the ora3-expressing (H) Enlarged dorsal view of the ora3-labeled cells (arrows) in the olfactory epithelia (delimited by the black dashed lines); (I) Enlarged dorsal view of the ora3-labeled cells (arrows) in the mouth region. (J) Lateral view shows expression of ora6 in the thymus and a thin layer of cells surrounding the yolk region (arrows); (K) Ventral view shows expression of ora6 in three pairs of branchial arches (arrows); (L) Dorsal view shows expression in the olfactory epithelia (arrows); (M) Enlarged view of one olfactory epithelia (circled) shows expression of ora6 in sparse olfactory sensory neurons; (N) Enlarged frontal view of the ora6-expressing mouth region. These experiments were kindly performed by Yen Yen Kwan (ora1 and ora6) and Heven Midani (ora3). Figure is in the previous page.

ora6 (Figs. VII-10A and J), pectoral fins for ora1 and ora3 (Figs. VII-10B and F), cells surrounding neuromasts in the head region for ora3 and ora6 (Fig. VII-10G and data not shown) and in the lips or mouth for region ora3 and ora6 (Figs. VII-10I and N).

14. Monogenic rule of expression valid within the ora gene family

For several mammalian OR, V1R and V2R genes it has been shown that individual receptor neurons only express a single receptor from the corresponding gene family. I used double-immunofluorescence RNA in situ hybridization to compare the expression of *ora* genes (for details see MM section). Double labeling using permutations of *ora* probes belonging to the same gene pair showed that different *ora* genes are expressed in different subpopulations of OSNs (Figs. VII-11 and VII-12). In the two *ora* gene pairs analyzed, ora1-ora2 and ora3-ora4, ora1 was never expressed together with ora2 (Fig. VII-11A), and in no case were ora3 positive neurons found to be positive for ora4 (Fig. VII-11B). As expected, double labeling experiments using the same *ora* RNA probes labeled with biotin and DIG for one and the same *ora* gene, showed a total overlap, thus showing that the method used is working and reliable (Figs. VII-C and D).

Similarly, further double labeling experiments with several other combinations of *ora* probes belonging to different gene pairs, showed that different *ora* genes are always expressed in a different subset of OSNs (Fig. VII-12 and data not shown). Despite the fact that not all the possible combinations were analyzed, I hypothesize that each *ora* gene defines their own subpopulation of OSN in the olfactory epithelium, consistent with the previously reported monogenic expression of several families of olfactory receptors in mammalian and fish species (Sato, Miyasaka et al. 2005; Roppolo, Vollery et al. 2007).

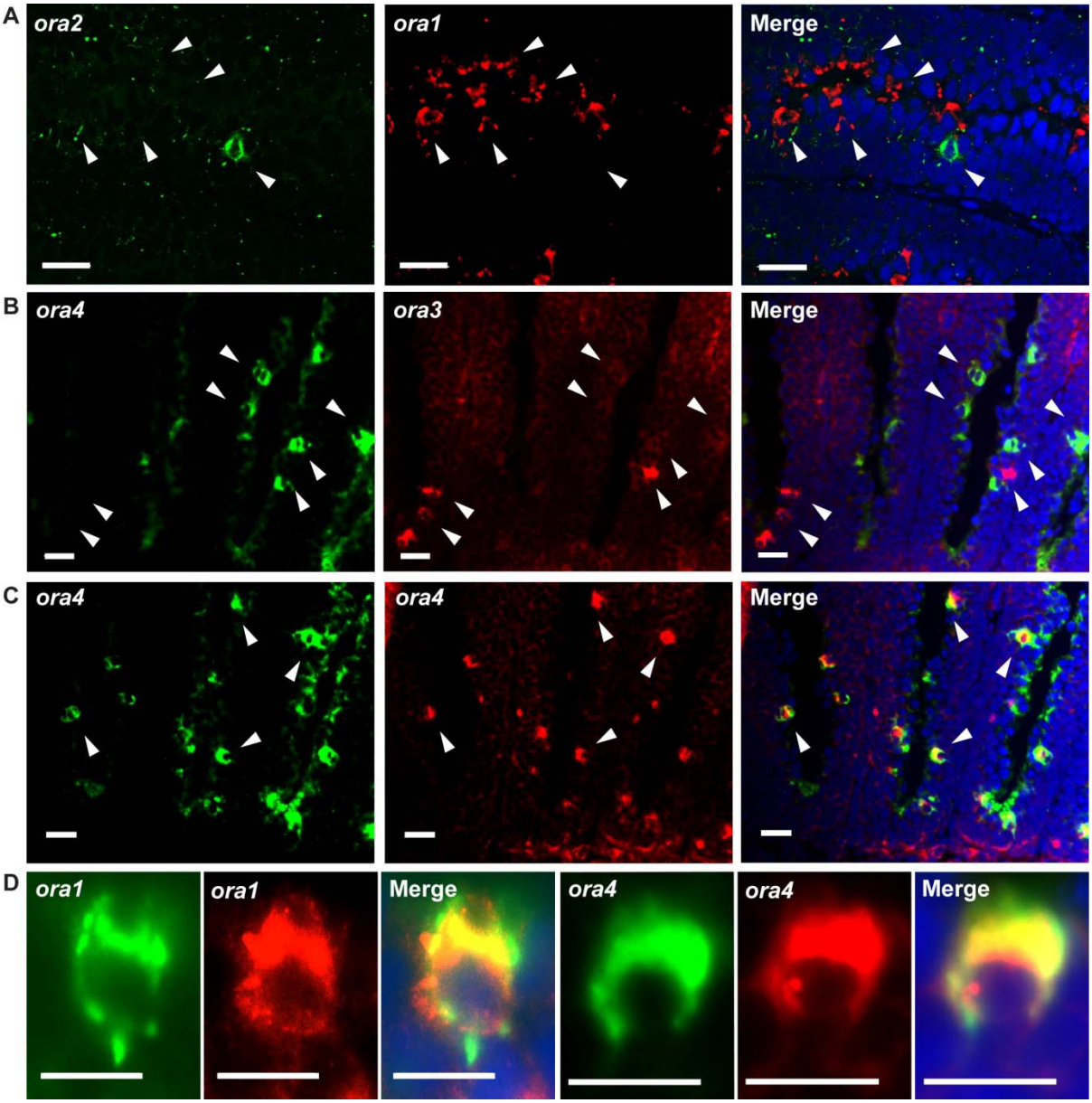


Figure VII-11 | ora genes do not co-express with members of the same clade.

(A-D) The olfactory epithelium expression patterns of ora genes belonging to the same clade were compared using two-colour RNA in situ hybridization. Probes for different ora genes labeled different OSNs (A, B), whereas probes for the same ora labeled the same cells (C). (D) An enlarged view of control experiments using the same ora probes, showing the co-label in the same olfactory sensory neurons. Scale bars, 20 um (A), 50 um (B-C) and 10um (D). Arrowheads point to the labeled neurons.

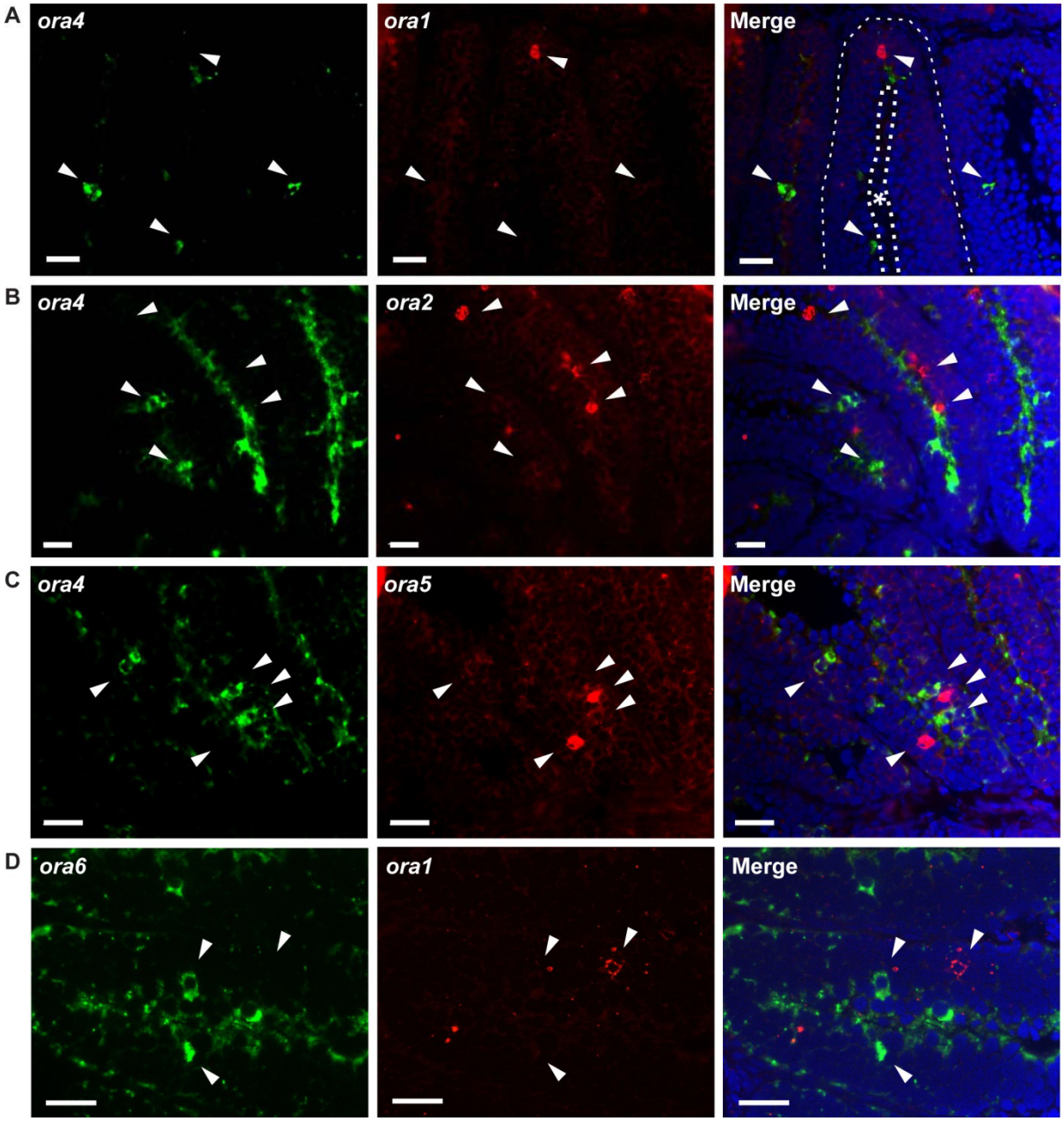


Figure VII-12 | *ora* genes of different clades do not co-express.

(A-D) The olfactory epithelium expression patterns of ora genes belonging to different clades compared using two-colour RNA in situ hybridization. Probes for different ora genes labeled different OSNs (A-D). Scale bars, 20 um (A-D). Arrowheads point to the labeled neurons.

B. TAAR family

1. TAAR genes are monophyletic and segregate from the monophyletic group of aminergic GPCRs

Mammalian trace amine associated receptors have recently been identified as another olfactory receptor gene family. In collaboration with Ashiq Hussain (cf Acknowledgments), I have identified and characterized the complete *taar* gene repertoire of five teleost fish, a shark, frog, chicken, four placental and one marsupial mammalian species was retrieved. All *taar* genes (from all species) analysed fell into 28 different subfamilies, indicated by numbers 1-28 (cf. Hussain *et al* 2008). Subfamilies 1 to 9 correspond to previously identified TAARs, with mostly mammalian members, whereas subfamilies 10 to 28 are fish-specific. The subfamilies segregate into three major classes, each of them representing a monophyletic clade (Fig. VII-11). Class I (TAAR1, 10-11, 21, 27) contains mostly teleost genes, class II (TAAR2-9, 12-13) comprises mostly tetrapod genes, and class III is restricted to teleosts (TAAR14-20, 22-26, 28). All *taar* genes identified, form a monophyletic group, clearly distinct from their closest relatives, the aminergic neurotransmitter receptors (Fig. VII-9). The TAAR gene family also segregates with maximal bootstrap values from the odorant receptors, which are less closely related, but belong to the same class of GPCRs, the rhodopsin type or class A GPCRs (Fredriksson, Lagerstrom et al. 2003). A group of lamprey GPCRs considered TAAR, in a recent publication (Hashiguchi & Nishida 2008), clearly clusters with aminergic receptors and not with TAARs (*Fig.* 13 and data not shown).

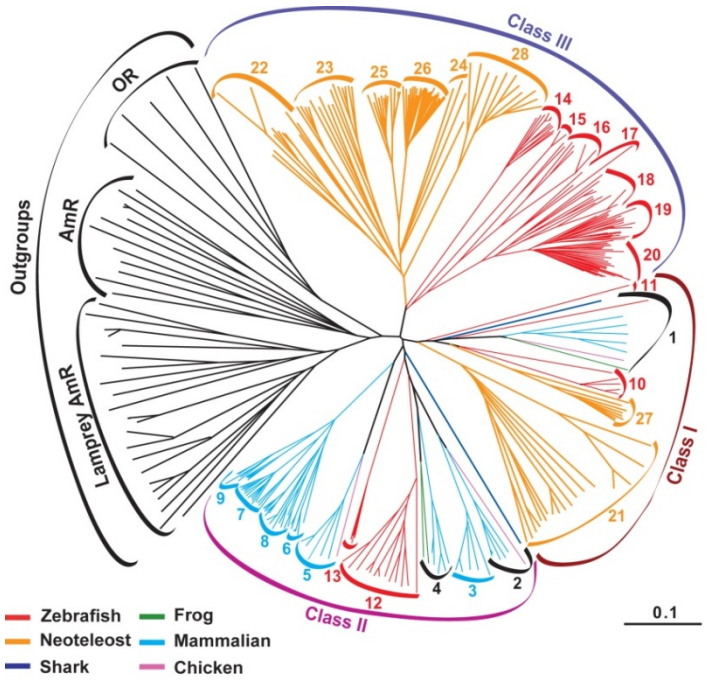


Figure VII-13 | Phylogenetic tree of TAAR family members.

Radial tree of teleost and tetrapod TAARs, species groups are color-coded. Zebrafish and mouse aminergic receptors were used as outgroup, together with a selection of odorant receptors (ORs). A group of lamprey receptors considered TAARs by (Hashiguchi and Nishida 2007) has aminergic receptors as closest neighbors, not TAARs. Scale bar, 10% divergence. For accession numbers and/or gene IDs see Figure 32 and Hussain et al 2008. Ashiq Hussain helped establishing the zebrafish taar repertoire and kindly provided all the other sequences included in this tree. Figure is in the previous page.

Despite overall heterogeneity, the TAAR family is characterized by distinctive consensus motifs t*aar* genes frequently show low identity values below 30% in pairwise comparisons (data not shown). We have therefore analyzed the retention of consensus motifs to obtain a second line of evidence for proper delineation of the *taar* gene family. The vast majority (41/48) of all amino acid positions reported as absolutely conserved between human and rodent TAARs (Lindemann and Hoener 2005) are highly to absolutely conserved in fish TAARs. These contain general GPCR motifs, but also many TAAR-specific motifs that are not found even in the closely related aminergic receptors (Fig. VII-14). The characteristic TAAR fingerprint motif, described to be 100% sensitive and specific for mammalian TAARs (Lindemann and Hoener 2005), is strikingly conserved in all fish *taar* genes analyzed (Fig. VII-14). In contrast, two of the TAAR-specific amino acids are absent in the lamprey receptors, and two others are only weakly conserved, further delineating the TAAR receptors from the group of aminergic receptors in general and from the lamprey receptor family in particular. As expected, amphibian, and avian TAAR sequences share the great majority of conserved motifs as outlined above, supporting their assignment as *taar* genes (data not shown). Some motifs distinguish the three classes of TAARs from one another (data not shown), including the aminergic ligand motif (Huang 2003), which is highly conserved in class I and II, but absent from all but one class III *taar* genes (Fig. VII-14).

2. Species-specific expansions of individual TAAR genes into subfamilies is a recurrent theme in all five teleost species examined

The teleost *taar* gene repertoires range from 112 for zebrafish (plus 4 pseudogenes) down to several fold smaller repertoires (stickleback 48, medaka 25, pufferfish each 18 genes). Mammalian families just reach minimal fish family size, while avian and amphibian families are minuscule, with only 3 genes each (*cf*. Hussain *et al* 2008). Most of these differences are caused by massive recent gene expansions in teleosts that led to 30 members within a single zebrafish-specific subfamily, TAAR20, and 28 genes in the stickleback-specific subfamily TAAR26. All but one zebrafish and one neoteleost TAAR subfamily (DrTAAR11 and TAAR24, respectively) have undergone recent gene duplications. In mammals gene expansions are less frequent and also much smaller those in teleosts

(maximally to six genes, opossum *taar9*). No recent gene expansions were found for *taar* genes 1, 2, 3 and 5. No recent gene duplications have been observed in an amphibian (*Xenopus tropicalis*) nor in an avian species (*Gallus gallus*). Individual teleost TAAR genes (except TAAR1) rarely possess any direct orthologs. Thirteen of nineteen subfamilies are restricted to a single species each (TAAR10-20, zebrafish; TAAR25-26, stickleback). Only two subfamilies contain genes from all four neoteleost species examined (TAAR21-22) and none are shared between zebrafish and neoteleosts, *cf*. SI Hussain *et al* 2008. Even in the case of subfamilies containing orthologs, a gene expansion may occur in one species but not another, e.g. TAAR27 has expanded to seven genes in tetraodon, but remains a single gene in both stickleback and fugu (*cf*. Hussain *et al* 2008). Thus most gene duplications have occurred rather recently, after the divergence of the teleost and neoteleost species analyzed here (Fig. VII-15), indeed even after the two pufferfish species diverged about 20-30 million years ago (Van de Peer 2004).

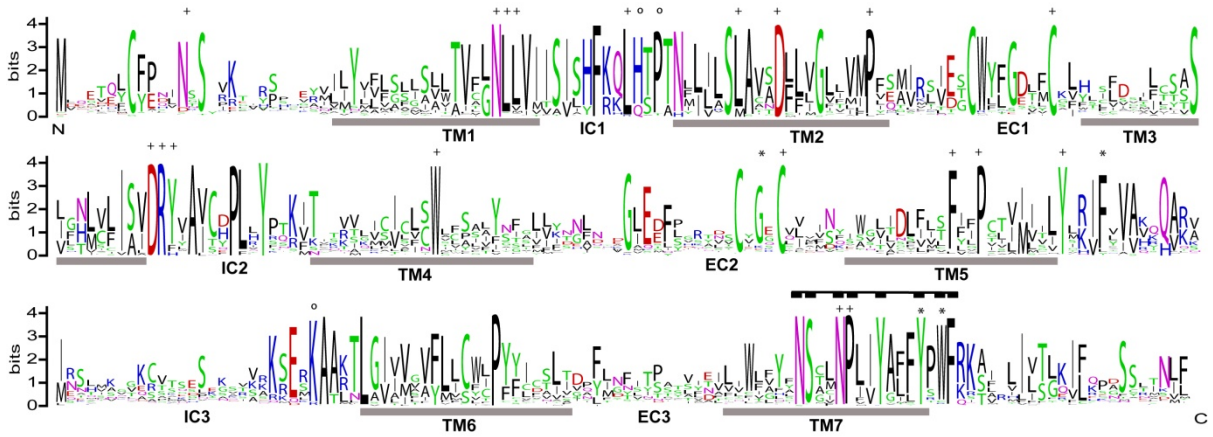


Figure VII-14 | Amino acid sequence conservation in the fish taar gene repertoire.

Sequence logo representation of the alignment of all 223 fish full length TAAR sequences, the height of the one-letter amino acid code in the logo reflects degree of conservation. Sequence logos were generated as described (Kraemer, Saraiva et al. 2008). TM, transmembrane region; IC, intracellular loop; EC, extracellular loop; plus signs, broadly conserved in rhodopsin type GPCRs; circles, conserved in some rhodopsin type GPCRs but not in aminergic receptors; asterisks, conserved in TAARs, but not in other rhodopsin type GPCRs. Two triangles, aminergic ligand motif (Huang 2003); filled rectangle motif in TM VII, characteristic fingerprint motif for TAARs, (Lindemann and Hoener 2005).

In contrast, orthologs are readily identifiable between all mammalian species analyzed. We uncovered orthologs for all nine previously identified mammalian *taar* subfamilies in another mammalian species, *Bos taurus* (*cf.* Hussain *et al* 2008). The missing subfamilies TAAR4 and TAAR7 are represented by pseudogenes. Thus, *Homo sapiens* has a typical mammalian TAAR repertoire. Seven of the nine subfamilies (TAAR1-6, 9) are detected also in opossum, a marsupial mammal, i.e. should be present already in the MRCA

of marsupials (Murphy, Pringle et al. 2007) and modern mammals. Although very small, the amphibian and avian *taar* gene repertoires are not located at the base of the sarcopterygian tree and clearly belong to different mammalian subfamilies. Thus gene losses appear to have shaped the avian and amphibian gene families.

3. Taar genes originate in the common ancestor of cartilaginous and bony fish

TAAR1 orthologs occur in both tetrapods and teleosts (Fig. VII-13), i.e. TAAR1 is evolutionary older than the actinopterygian/sarcopterygian split. To obtain a better understanding of the evolutionary origin of the *taar* gene family we have searched all currently available sequence information for cartilaginous fish and jawless fish. Two *taar* genes, both with a perfectly conserved TAAR-specific fingerprint motif (Lindemann and Hoener 2005), were uncovered in the elephant shark, one of them an ortholog of TAAR1 (Fig. VII-13). Cartilaginous fish are considered basal to all jawed vertebrates (Venkatesh, Erdmann et al. 2001), so TAAR1 was present already in the MRCA of bony fish and cartilaginous fish and may be the ancestral member of class I. All tetrapod species analyzed contain a TAAR1 ortholog, as does the avian genome examined here. Interestingly, no orthologs for TAAR1 could be found in any of the *neoteleost* species analyzed, i.e. this ancestral gene appears to have been lost in neoteleosts. The other shark gene exhibits a basal location in class II (Fig. VII-13) and may thus correspond most to the ancestral class II *taar* gene. No shark representative of the teleost-specific class III was found, consistent with a later evolutionary origin of this class, after the segregation of teleost and tetrapod lineages. No *taar* genes were uncovered in the genome of a jawless vertebrate (*Petromyzon marinus*). Thus the *taar* gene family appears to have originated in the MRCA of cartilaginous and bony fish as a pair of genes that later expanded into class I and II.

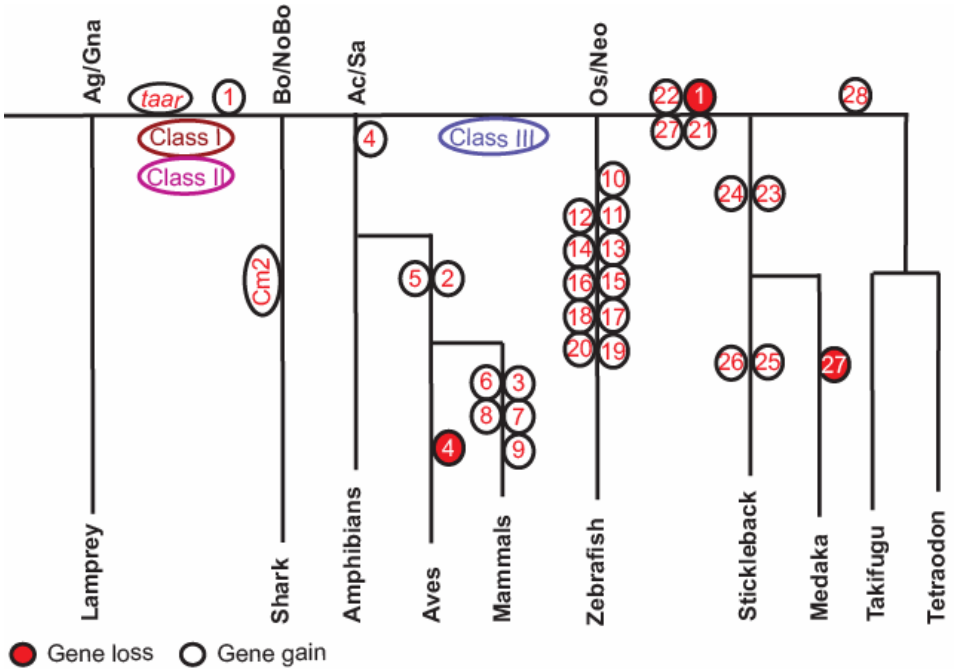


Figure VII-15 | Estimated minimal evolutionary age of TAAR subfamilies and genes..

Open circles represent the gene gain events in each lineage, and black circles represent the gene loss events. Inside each circle is the name of the respective gene(s) or subfamilies. Emergence of the taar gene family and of the three classes of taar genes is indicated by ovals. The major phylogenetic transitions are indicated: bo/nobo, bony fish/cartilaginous fish; ac/sa, actinopterygian/sarcopterygian split; os/neo, ostariophysii/neoteleostei segregation. The maximum parsimony principle was followed, thus gene gains are depicted at the last possible stage before additional gains would become necessary for explanation, but may in fact have occurred earlier. Picture is in the previous page.

4. Signatures of positive selection in teleost taar genes are masked by global negative selection

The phylogenetic analysis of the *taar* gene family dynamics suggests a faster evolutionary rate in teleosts than in tetrapods. To better understand both the variability and evolutionary dynamics of the *taar* genes we analyzed the selective pressure on these genes using both global and local analysis of substitution rates in synonymous *vs*. non-synonymous base positions. The global dN/dS values calculated for each of the ortholog groups show that all the gene groups are under negative selection (Fig. VII-16, *cf.* Hussain *et al* 2008), but the extent varies considerably, from 0,09 (pronounced negative selection) up to 0,8 (close to neutral selection). The average dN/dS value for the teleost-restricted class III is by far the highest, more than double the value for class II *taar* genes and significantly different from both class I and class II values (Fig. VII-16). The relaxed negative selection observed for class III TAAR subfamilies may result from an overall pronounced negative selection masking positive selection at some sites. *In* order to clarify this point, I analyzed for all subfamilies the dN/dS values for each individual codon position for all genes of each TAAR subfamily. As predicted by the analysis of the previously calculated global dN/dS values, negatively selected sites were found without exception throughout all the *taar* gene families, with some preponderance in the transmembrane regions (Fig. VII-14). Consistent with the results of the global analysis, class III *taar* genes contain only about half as many negatively selected sites as the other two classes (*cf.* Hussain *et al* 2008).

More excitingly, the site-by-site dN/dS analysis revealed a significant number of positively selected sites, which were masked by the preponderance of negative selection in the global analysis. While there are very few such sites in class I and II *taar* genes (0 to 2 sites per gene), several genes in class III show much higher values of up to 20 sites per gene (Fig. VII-16, *cf.* Hussain *et al* 2008). We conclude that the teleost-restricted class III, which is evolutionary much younger than class I and class II, has undergone more rapid evolution resulting in massive expansion of gene families beyond that observed in the older classes I and II.

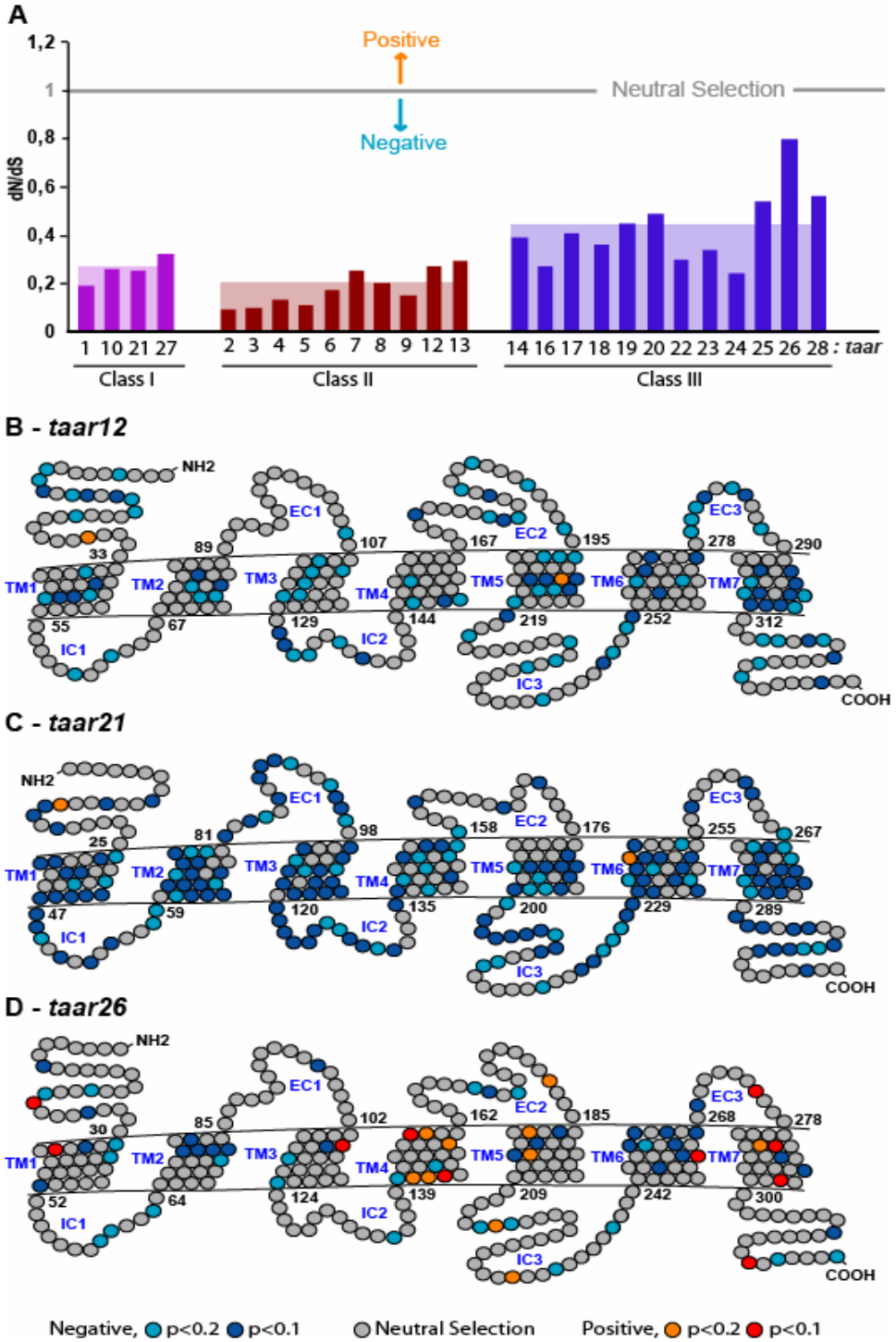


Figure VII-16 | Evolutionary distances and selective pressure on taar genes.

(A) dN/dS ratios of the TAAR ortholog groups in which this analysis was possible (>2 genes/group). Genes are arranged by class, the class average is indicated by background shading. (B-D) A representation of site-by-site selective pressure is shown for three TAAR sequences (negative selection in light-blue, p<0.2 or blue, p<0.1, neutral selection in gray, positive selection in orange, p<0.2, and red, p<0.1). (B) Results for zebrafish-specific TAAR12, a class II subfamily. (C) Results for TAAR21, a class I subfamily, which includes ortholog genes of all four neoteleost species. (D) Results for stickleback-specific TAAR26, a class III subfamily.

5. Teleost taar genes exhibit more diverse genomic location than tetrapod and avian taars

Mammalian *taar* genes are found without exception in a single cluster in the genome (Lindemann, Ebeling et al. 2005). All newly identified mammalian, avian and amphibian *taar* genes conform to this previously described pattern. In contrast, teleost *taar* genes are found in two large clusters and a few solitary genes (chromosomal allocation for zebrafish and medaka, large scaffolds for stickleback). Within the clusters, genes are organized mostly in accordance to phylogenetic relationship (Fig. VII-13), consistent with a genesis of the clusters by recurrent local gene duplication. A few exceptions to the co linearity of phylogenetic relationship and genomic location do occur (*cf.* Hussain *et al* 2008), possibly caused by recent genomic rearrangements involving these genes.

Interestingly, *taar*1 gene is always located at one end of the cluster in tetrapod and avian species, consistent with an asymmetric process being responsible for at least some of the repeated gene duplications. In contrast, zebrafish *taar*1 is situated in the middle of a large gene cluster.

Average intergenic distance is 7.9+/-0.5 kb (mean+/-SEM, n=97) in the zebrafish gene clusters, with exception of a large intervening region at approximately the same relative position in both clusters (*cf.* Hussain *et al* 2008)**.** This similarity in cluster structure is consistent with the two clusters resulting from the whole genome duplication in early teleosts (Nakatani, Takeda et al. 2007). Indeed, the cluster positions for zebrafish and medaka are syntenic not only within and between species, but also to the human cluster (*cf.* Hussain *et al* 2008, Woods *et al*, 2005; Nakatani, Takeda *et al* 2007).

Class III *taar* genes are found in both genomic clusters, whereas nearly all class I and class II genes are restricted to one (and the same) cluster. Consequently, even class III appears to be older than the whole genome duplication observed in early teleosts (Nakatani, Takeda et al. 2007). Since on the other hand class III is restricted to teleosts, it appears to have originated shortly after the segregation of the teleost and tetrapod lineages.

6. Most taar genes are expressed in sparse olfactory sensory neurons

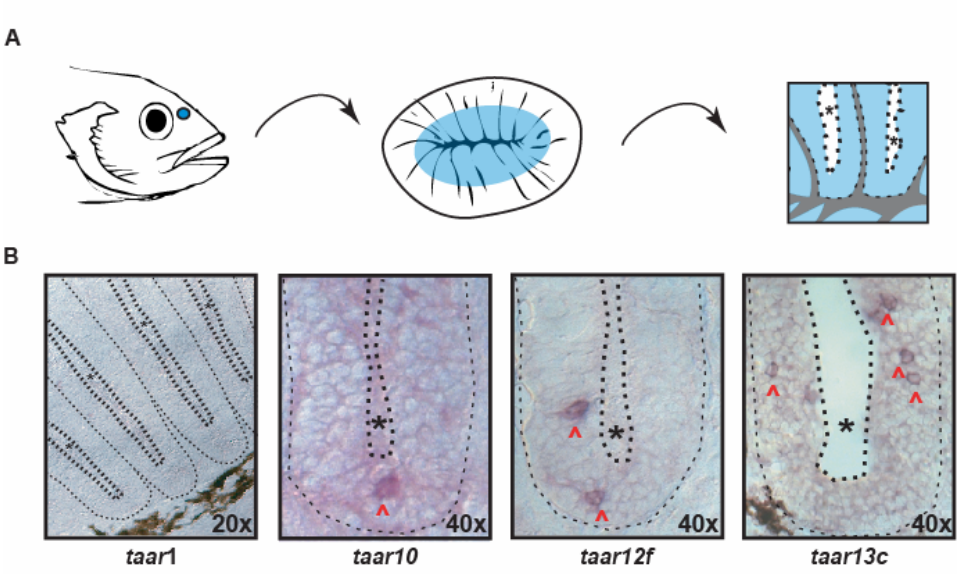
The rapid evolution and positive selection observed in the *taar* gene family in teleosts are consistent with expectations for olfactory receptor genes (*cf*. Alioto and Ngai 2005), since efficient adaptation to changing environmental odor stimuli may require high evolutionary rates. Another essential requirement for olfactory receptor genes is an expression in the olfactory epithelium. A representative subset of four *taar* genes (class I, TAAR1, 10; class II, 12f, 13c) was chosen for analysis of expression. TAAR1 was included

for its special phylogenetic position. Probes for *in situ* hybridization were chosen to minimize cross-reactivity with related *taar* genes as far as possible.

Except TAAR1, all genes tested were expressed in the adult zebrafish olfactory epithelium (Fig. VII-17). Labeled cells were sparsely distributed within the sensory area of the olfactory epithelium. No expression was observed in the outer, non-sensory ring of the nasal epithelium. This labeling pattern is reminiscent of that observed for other zebrafish olfactory receptors (Weth, Nadler et al. 1996; Saraiva and Korsching 2007; *cf*. Sato, Miyasaka et al. 2005) and thus supports an expression in olfactory sensory neurons, consistent with an expression of most or all *taar* genes in these neurons.

Figure VII-17 | Expression of taar genes in the zebrafish olfactory epithelium (OE).

(A) Schematic representation of the localization of the OE followed by a drawing of a horizontal section of OE (lamellae are cut perpendicular to their flat face) and finally an enlargement of two lamellae. The central blue-colored area in the lamellae indicates the location of the sensory neuroepithelium, cf. (Weth, Nadler et al. 1996), gray areas and thin dotted line, basal lamina; black dots and asterisk, lumen. (B) In situ hybridization in horizontal sections of the OE with antisense RNA probes. The area shown corresponds roughly to one half of the schematic representation in the right panel of A). Red arrowheads point to labeled neurons, other symbols as in A).



C. Molecular characterization of OSN subtypes with respect to receptors, G-proteins and s100 calcium binding proteins

To date no olfactory receptor gene family was associated with the most recently discovered OSNs, the crypt cells. Likewise, the cellular properties of the *ora* genes are yet to be characterized. The spatial pattern depicted by *ora* genes resembles the one described to the crypt cells. To test this assumption I characterized the S100 antibody immunostaining pattern in zebrafish and performed double-labeling experiments of S100 antibody and *ora* genes.

1. Crypt cells are detected in the apical region of the lamella of adult zebrafish olfactory neuroepithelium

A commercially available antibody against bovine S100 family members S100B and S100A1 has previously been shown to label morphologically recognized crypt cells in the adult olfactory epithelium of zebrafish (Germana, Montalbano et al. 2004) and some other teleost fish species (Schmachtenberg 2006, Hamdani and Doving, 2005) as well as skate, a cartilaginous fish (Ferrando *et al* 2007). Crypt cells are described to be a rare cell population, present in very small numbers in all the fish species described (refs Hansen et al JN 2003). I decided to test this assumption by analyzing the S100 immunoreactivity in the adult zebrafish olfactory neuroepithelium, in particular the frequency and spatial distribution of S100+ cells. In adult epithelia many cells with distinctly neuronal morphology are labeled by the S100-antibody (Fig. VII-18A). Their oval cell body shape and apical distribution in the epithelium are consistent with results obtained by Germana using the same antibody (Germana, Montalbano et al. 2004) and others using other labeling techniques (Schmachtenberg, 2006). Surprisingly, a high density of labeled cells was found dispersed in the apical region of the lamellae of the adult zebrafish olfactory epithelium, in contrast to the previously reported low numbers and density for this neuronal cell type (Fig. VII-18A). Moreover, S100 immunoreactive cells (i.e. crypt cell neurons) can display either a scattered isolated distribution pattern (Fig. VII-18B) or occur in crypt cell clusters (Figs. VII-18C-D), the later more commonly found in the dorsal region of the olfactory epithelium (Fig. VII-18E).

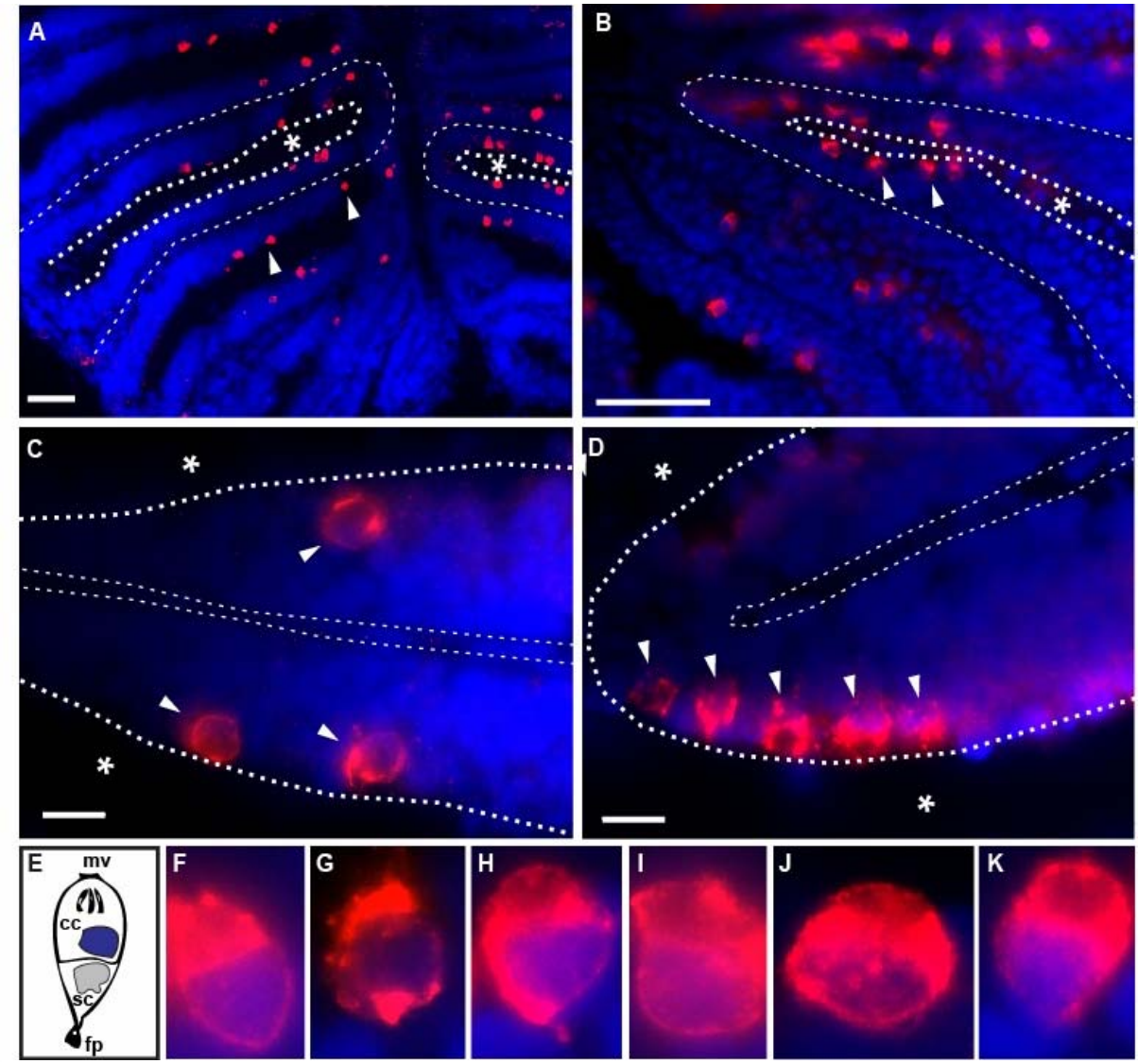


Figure VII-18 | S100-immunoreactivity in the adult zebrafish olfactory epithelia.

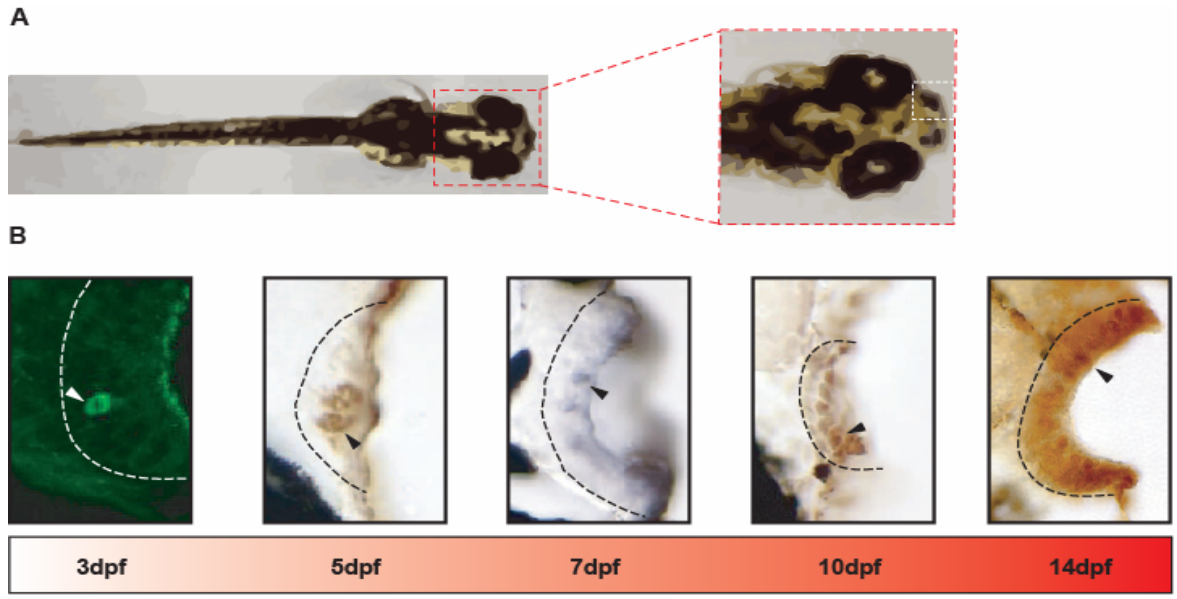
(A-D) Immunoreactivity (IR) of S100 in the adult olfactory epithelium show that crypt cells mainly locate in the apical region of the lamella. Crypt cells can show a sparse distribution across the lamellae (C) or be aggregated in clusters (D). Thin dotted line, basal lamina; black dots and asterisk, lumen. E) schematic representation of a crypt cell. Mv, microvilli; cc, crypt cell; sc, supporting cell and fp, foot process. (F-K) Morphological variety displayed by crypt cells: crypt cells can display a more elongated shape (F, G) or more globose morphology (I-K). Scale bars, 50 um (A,B) and 10 um (C-K).

2. Crypt cells are detectable by immunoreactivity already three days post fertilization

Previously crypt cells had not been found before juvenile developmental stages. Consequently it has been suggested that crypt cells may originate or differentiate only during sexual maturation of the fish (Yamamoto, Mori et al. 2004). I tested this assumption by examining the developmental time course of S100 immunoreactivity in zebrafish. An elevated density of labeled cells was found in juvenile zebrafish at an age of 14 days post fertilization, where many of the cells in the developing epithelium seem to be reactive (Fig. VII-19). In the developing olfactory placode of 10 days post fertilization (dpf) old fish the number of labeled cells is still high (Fig. VII-19). At the age of seven days post fertilization less cells were labeled, but a few single cells in the olfactory placode are clearly visible (Fig. VII-19), as well as in five days post fertilization old larvae (Fig. VII-19). The youngest fish analysed (three days post fertilization) showed immunoreactivity as well (Fig. VII-19). Since at this stage very few cells were labeled, day three post fertilization appears to be close to the onset of either crypt cell genesis or S100 expression. In all developmental stages tested, labeled cells are large and round (cf. Fig. VII-19), consistent with the known morphology of mature crypt cells (Hansen and Zeiske 1998; Hansen and Zielinski 2005).

Figure VII-19 | S100-immunoreactivity in larvae-juvenile zebrafish olfactory epithelia.

(A) Schematic diagram of a typical zebrafish larvae showing the location of the olfactory epithelium (white dashed square). (B) S100-IR in 14 days post fertilization old zebrafish reveals strong signals. Ten days post fertilization old olfactory epithelium with S100-IR. Three marked crypt cells in seven days post fertilization and five days post fertilization old epithelium. The youngest stage examined with three days post fertilization old fish revealed S100-IR as well. Olfactory epithelium is encircled by the black dashed lines and arrows point at the labeled crypt neurons. Andreas Kraemer kindly performed the experiments.



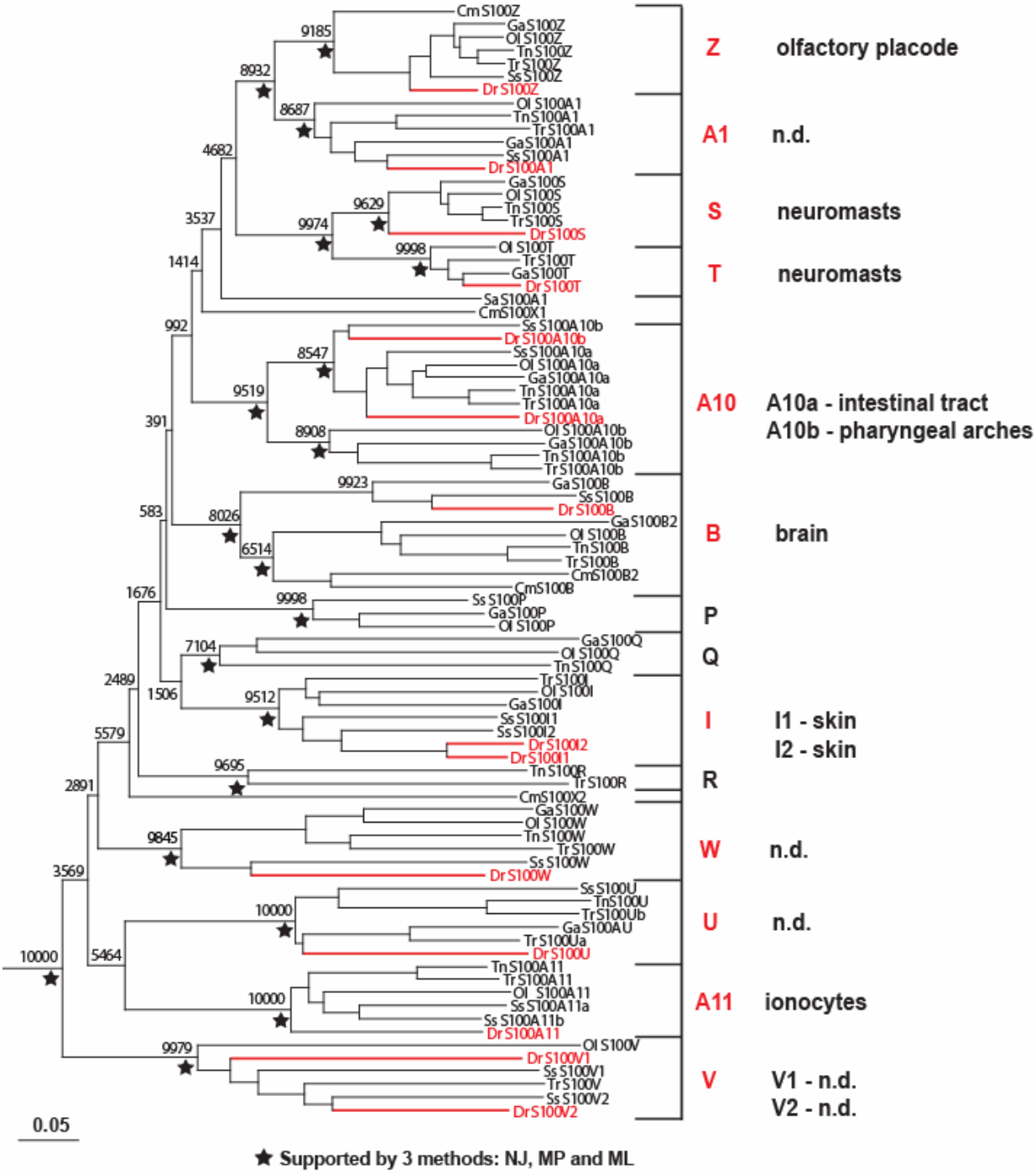


Figure VII-20 | S100 family members in zebrafish.

Phylogenetic tree and main site of expression (cf. Kraemer, Saraiva et al. 2008) are shown. Fish s100 genes from six teleost species and two cartilaginous fish (*Squalus acanthias* and *Callorhinchus milii*) are depicted. The red colored names and branched indicate zebrafish s100 genes. Stars indicate that the clades downstream of the node are supported by all three methods used for the phylogenetic analysis (NJ, ML and MP). The tree presented was constructed using the NJ method. Bootstrap support (total 10000 replications) is indicated at the major nodes. Scale bar indicates the number of amino acid substitution per site.

3. S100Z is specifically expressed in the olfactory placode

It is not known, which gene is responsible for the observed cross-reactivity to the S100 antibody in the zebrafish olfactory epithelium. Under the assumption that it is indeed a member of the S100 gene family, it should be identifiable by *in situ* hybridization screening, once the complete repertoire of zebrafish S100 genes is known. Working together with Andreas Kraemer (cf Acknowledgements), I have recently characterized the S100 gene family in teleost fish (Kraemer, Saraiva et al. 2008), which contains fourteen members, some of them without mammalian orthologs (Fig. VII-20). An overview of the larval expression pattern of ten zebrafish S100 genes showed two genes with an expression in the olfactory epithelium (Kraemer, Saraiva et al. 2008), one of which, S100Z was detected exclusively in this organ (Figs. VII-20 and VII-21). Incidentally, S100B and S100A1, the genes responsible for the antibody labeling in mammals, while present in zebrafish, are not expressed in larval stages (Kraemer, Saraiva et al. 2008). Thus, so far S100Z appears to be the most likely candidate for the observed antibody labeling. Due to time constraints, the remaining four genes have not been analyzed yet. The developmental time course of expression of S100Z was studied, using *in situ* hybridization with sense and antisense probes and several stages during embryonic and larval development. S100 expression is always restricted to the olfactory epithelium (Figs. VII-21A, B). At day three post fertilization a handful of cells are labeled (Fig. VII-21B), whose globose morphology (Fig. VII-19) is consistent with an expression of s100z in crypt neurons. Already at day two weak expression is found in several cells of the olfactory placode (Fig. VII-21D), with occasional strongly labeled cells. Thus the onset of S100Z expression appears to be around day 2 to day 3, consistent with the onset of the S100 immunoreactive crypt cells.

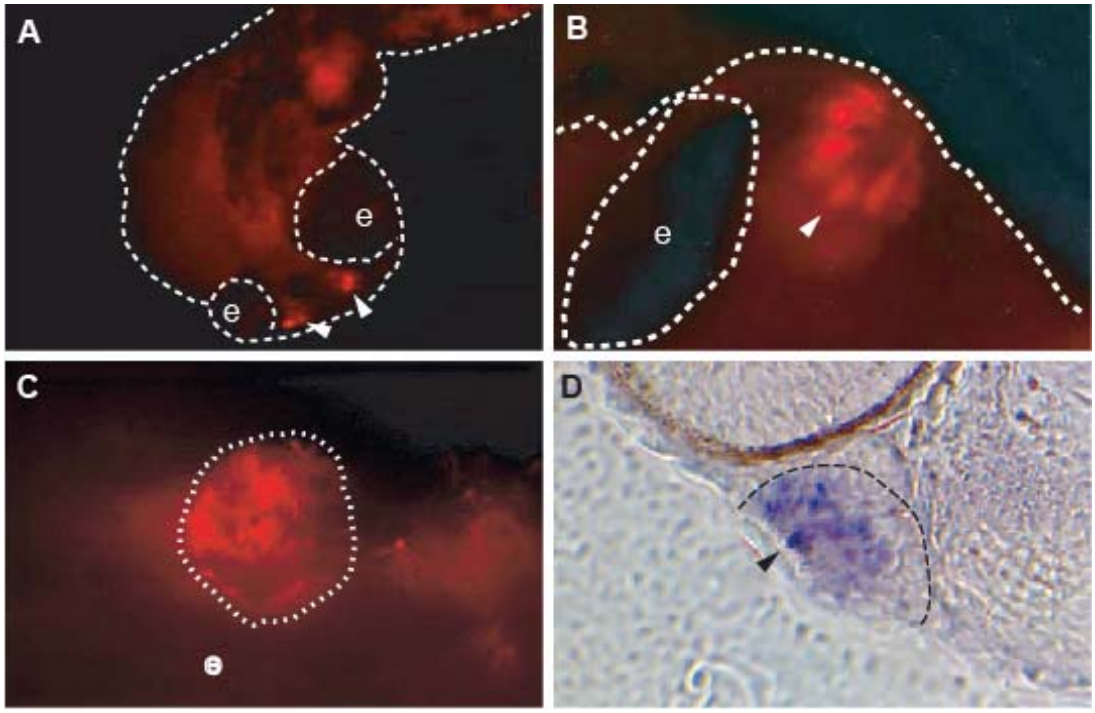


Figure VII-21 | Developmental time course of S100z expression as detected by in situ hybridization.

(A) and (B) S100Z in situ hybridization indicates specific expression in the olfactory placode of zebrafish larvae three days post fertilization. The sole signal detected in whole mount in situ hybridization is restricted to the olfactory placode. The weak diffuse fluorescence in head and trunk is unspecific, as it also occurs in the sense control (data not shown). (C) Cryosections of the same fish shown in A) and B) reveal large single cells with crypt-like morphology. (D) Cryosection of an 48 hpf old embryo shows strong expression in one neuron (arrow) and several weakly labeled cells in the olfactory placode. Andreas Kraemer kindly performed the experiments and the pictures to me.

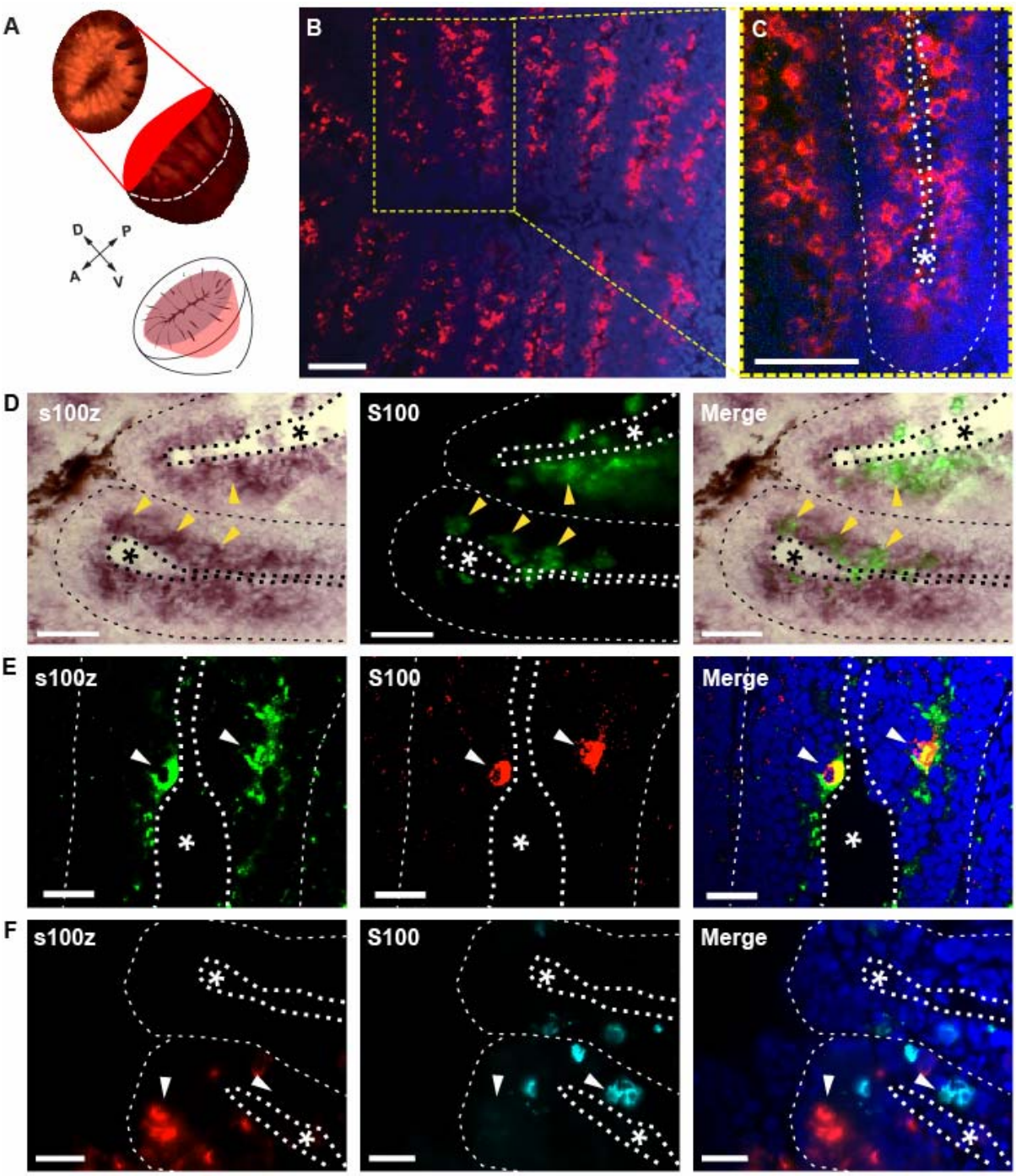
4. Spatial pattern of s100z and S100 antibody labeling are markedly similar

In the horizontal dimension, S100Z-expressing cells are restricted to the inner sensory region of the epithelium (Fig. VII-22) (*cf*. (Weth, Nadler et al. 1996). Vertically, S100z expression is mostly limited to the upper, dorsal segment of the sensory region (Fig. VII-22A), as shown by whole mount *in situ* hybridization of the adult olfactory epithelium (Fig. VII-22A). *In situ* hybridization on cryosections (FigS. VII-22B and C) demonstrates expression of S100Z in a large population of neurons, which exhibit a similar apical position within the lamellae. However s100z expressing cells occur at a seemingly higher frequency when compared to the number of neurons labeled by the S100 antibody. This broad, but nevertheless restricted expression of S100Z is reminiscent of the broad, but spatially restricted expression domains described for other olfactory receptor gene markers (e.g. OMP and TRPC2) in the zebrafish (Weth, Nadler et al. 1996; Weth 2001) and is thus consistent with an expression of S100Z in a subpopulation of olfactory sensory neurons.

In an attempt to unequivocally decide whether s100z is responsible for the observed S100 antibody labeling of crypt cells, I performed simultaneous labeling of S100 protein and s100z RNA by immunostaining and *in situ* hybridization, respectively, using three different methods. In two out of the three methods, an overlay of the signals obtained by the s100z probe (NBT-BCIP and TSA Kit) and the S100 antibody revealed overlapping signals (Figs. VII-22D-F and G-I) of both detection methods at the cellular level. This results suggest that s100z causes the S100 immunolabeling of crypt cells, which has been established as a crypt cell marker (Germana, Montalbano et al. 2004), thus making s100z a suitable molecular marker for crypt neurons. It was therefore completely unexpected that preliminary experiments using a different fluorescent labeling detection for the s100z probe (HNPP-FastRed Kit) together with the S100 antibody revealed an exclusively non-overlapping expression of the labeled cells (Fig. VII-22J-L). This issue cannot be simply explained by a possible quenching of the green or red fluorescence by one or the other fluorophore, since control experiments using the same detection method with both probes for the same genes resulted in perfect overlap, as expected (see Fig. VII-11). While it cannot be excluded that one of the four genes not tested so far is responsible for the S100 antibody labeling of crypt neurons, the discrepancy between the three labeling methods prohibits the use of s100z as a crypt cell marker until this methodological ambiguity is resolved. In the subsequent experiments I therefore continue to use the S100 antibody as crypt cell marker.

Figure VII-22 | In situ hybridizations of s100z on adult olfactory epithelia and co-expression experiments with S100 antibody.

The S100Z expression is restricted in the horizontal dimension to the inner part (sensory area) of the olfactory epithelium (A) and vertically to the one first third of the dorsal part of the epithelium (B-C). The cryosection reveals single cells along the apical part of the lamellae (C-D). In two of the methods used for the co-labeling experiments (D, s100z NBT-BCIP + S100 AlexaFluor488; E, s100z HNPP-FastRed + S100 AlexaFluor488), s100z expressing neurons and S100 reactive crypt neurons reveal overlapping signals (D,E). Experiments using a different method (F, s100z TSA-Kit Alexa Fluor 488 + S100 AlexaFluor594) showed total exclusion of both labeling methods. Thin dotted line, basal lamina; black dots and asterisk, lumen. Arrows indicate labeled cells. Scale bar, 50 um (B, C) and 10 um (D-F). Figure is in the next page.



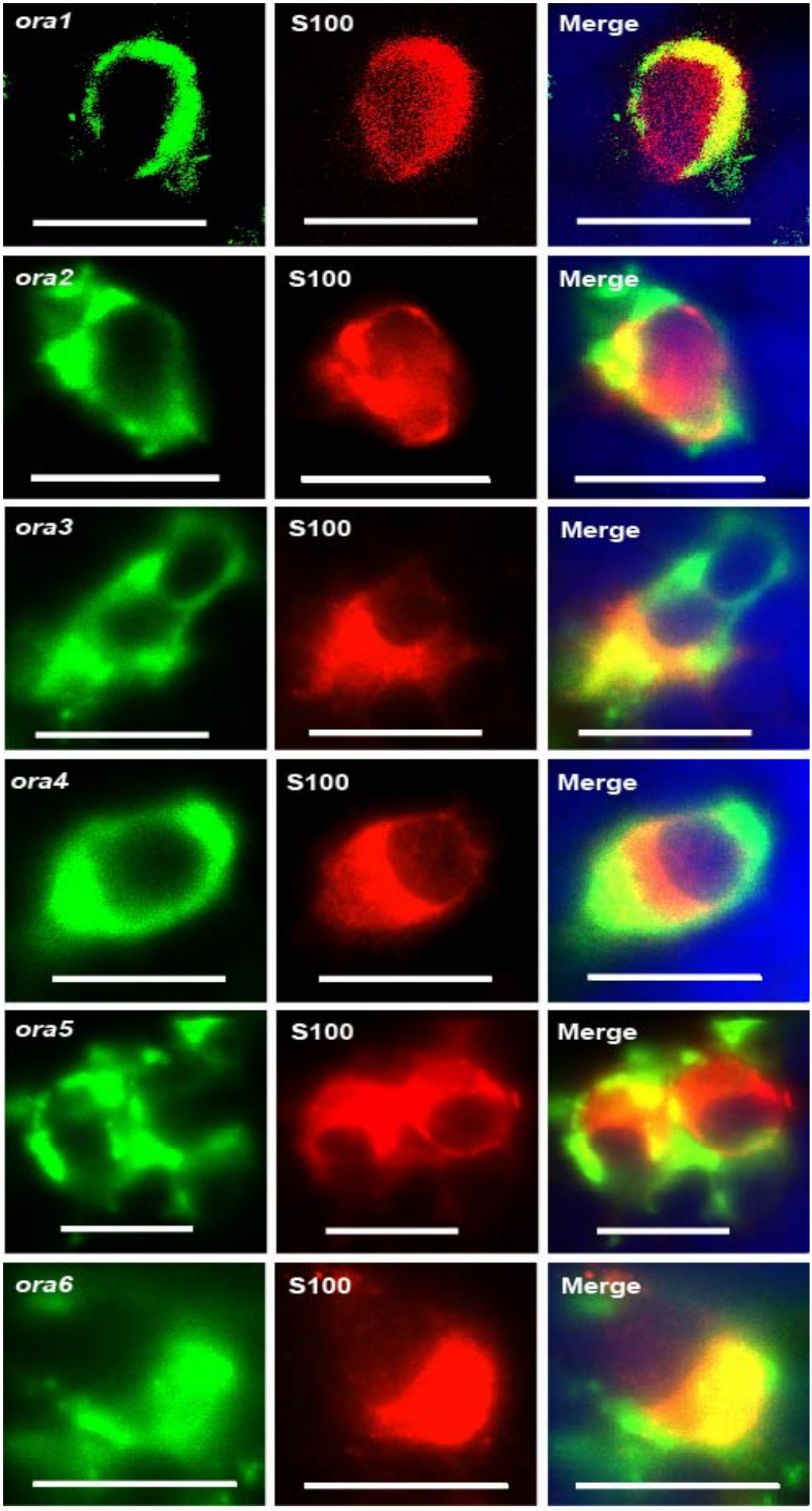
5. ora genes deorphanize crypt cell neurons

While the majority of the *ora*-expressing OSNs lie within the apical region of the lamella (Figs. VII-9, VII-10 and VII-11), the zebrafish OE contains three distinct populations of OSNs; crypt, microvillous and ciliated neurons, which are located in the apical, medial and basal regions of the lamella, respectively (Hansen and Finger 2000; Hansen, Rolen et al. 2003; Hansen, Anderson et al. 2004; Hansen and Schmidt 2004; Hamdani, Lastein et al. 2007). Making use of specific markers for each of the OSN populations, I examined by double

fluorescent in-situ hybridization, which of these three cell populations express the *ora* receptors. As predicted by its spatial distribution, all *ora* genes were found to be expressed in the S100-immunoreactive neurons (Fig. VII-23), which are specific crypt cell markers, thus deorphanizing this third class of olfactory sensory neurons. Additional double labeling experiments using mixed *ora* and the s100z probes revealed that *ora* genes do not co-express s100z (Fig. VII-24C-D). These results are in line with the exclusive expression of s100z and S100 antibody using the same detection method reported in the preceding paragraph. As a control I also performed double in-situ hybridizations of the *ora* genes with a previously described ciliated cell marker, as well as a microvillous cell marker. These results showed that *ora* genes are not expressed in ciliated OMP+ neurons (Fig. VII-24A-B). Accordingly to previously reported results, OMP probe labeled only ciliated neurons whose cell bodies are located in the basal layer of the olfactory epithelium (Fig. VII-24A). As marker for microvillous OSNs I chose OlfCa1 (Alioto and Ngai 2006) which our group has shown to label microvillous cells (Kumar, A and Weth, F, unpublished results), similarly to the established TRPC2, but with higher intensity. OlfCa1 is the ortholog of the goldfish 5.24 and previously named ZO6. It is broadly expresses in the medial region of the lamella (Fig. VII-25A-B), comparable to the expression pattern depicted by the TRPC2 gene (Sato et al 2005). The medial region of the lamella is known to be where the majority of the microvillous cells are located (Hansen and Finger 2000; Hansen, Rolen et al. 2003; Hansen, Anderson et al. 2004; Hansen and Schmidt 2004; Sato, Miyasaka et al. 2005; Hamdani, Lastein et al. 2007), therefore making it a suitable microvillous cell type marker. Moreover, this situation is reminiscent of its mammalian ortholog, the mouse V2R2, that labels the whole basal Gαo+ region of the vomeronasal organ (Matsunami and Buck 1997; Ryba and Tirindelli 1997; Martini, Silvotti et al. 2001). Double labeling experiments with OlfCa1 showed exclusive expression of the *ora* and OlfCa1 genes (Fig. VII-25A-B). These results are consistent with an exclusive expression of the *ora* receptors in the crypt cell neurons.

Figure VII-23 | Crypt cells expressing ora genes.

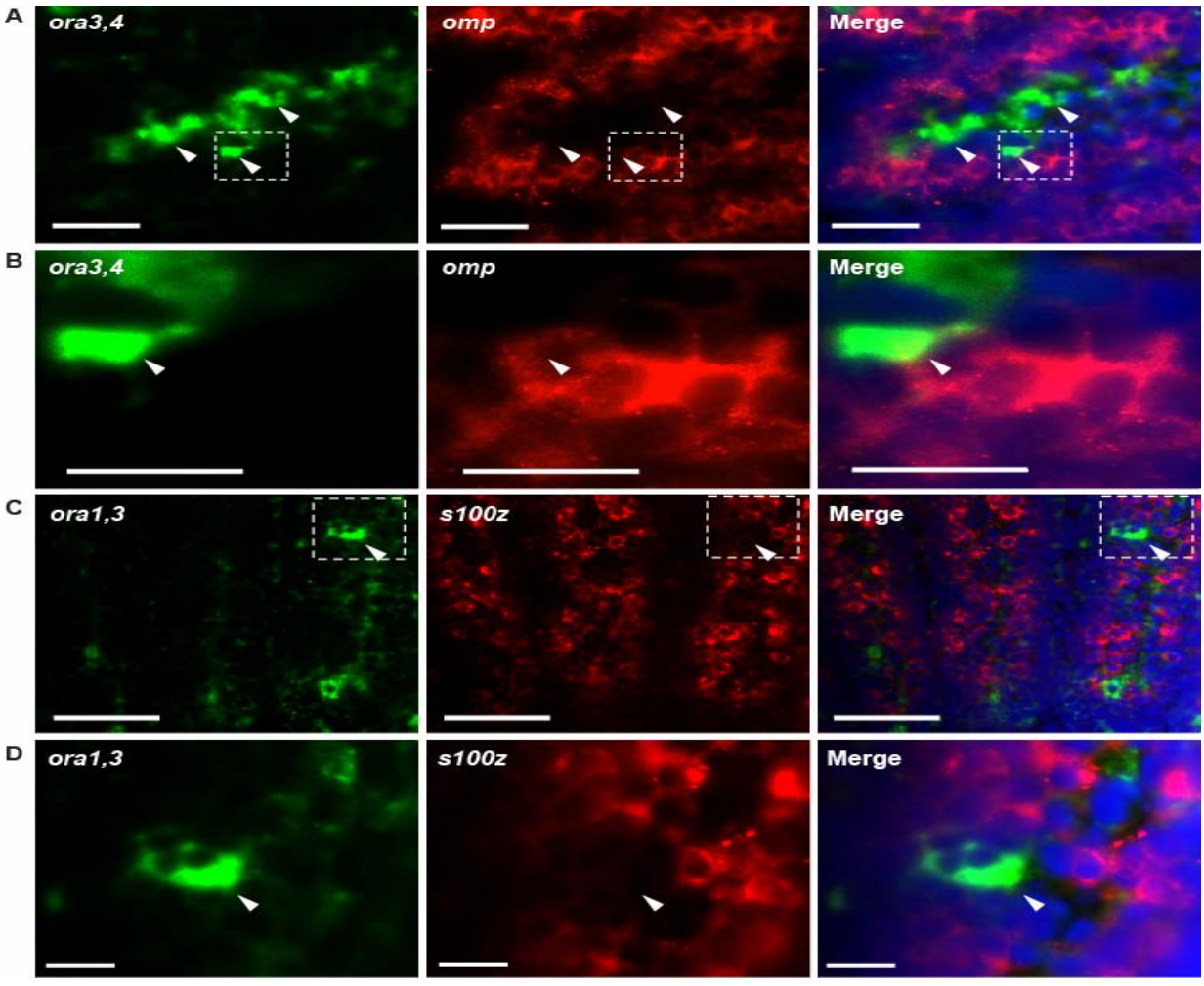
Detailed view of S100-immunoreactive crypt cells co-expressing ora genes. Double labeling experiments using a single ora probe together with the S100 antibody were performed. All ora genes were tested individually against the S100 antibody. These experiments show that all the ora-positive OSNs were found to be also labeled by S100 antibody. In contrast, not all the S100-positive crypt cells were labelled by a single ora probe (data not shown). Scale bar, 10 um. Figure is in the next page.

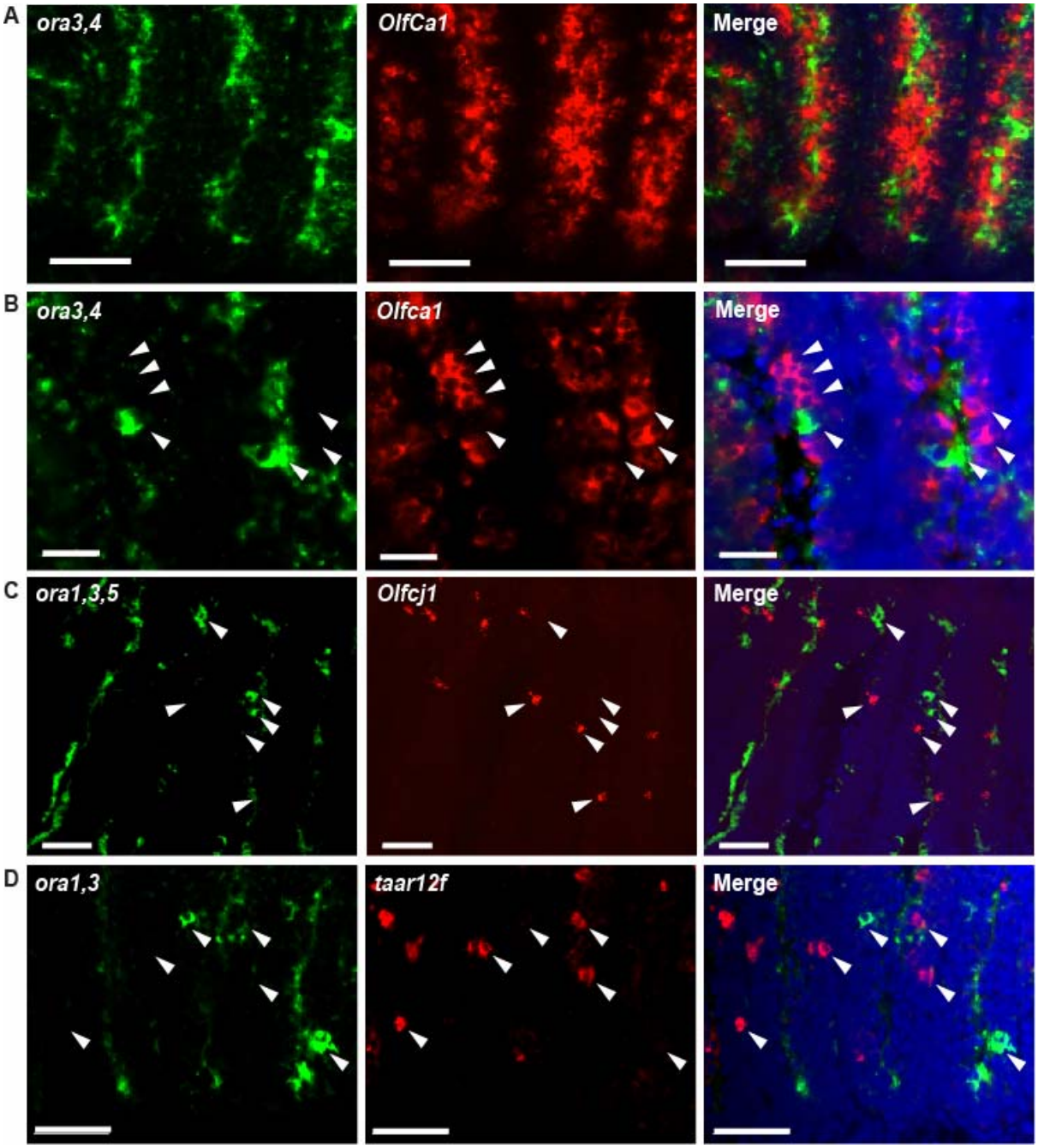

ora1
S100
Merge
ora2
S100
Merge
ora3
S100
Merge
ora4
S100
Merge
ora5
S100
Merge
ora6
S100
Merge

6. ora genes maintain exclusiveness towards other olfactory receptor gene families

First, the *ora* expression pattern was compared to that of two OlfC family members and as predicted by the result obtained with the OlfCa1 (Figs. VII-25A and B), *ora*-positive neurons are also negative for both OlfCj1 (Fig. VII-25C) and OlfCu1 (data not shown) (both are expressed in a seemingly higher number of neurons and primarily in the medial region of the lamella). Lastly, the *ora* genes were compared to two representatives of the taar gene family with respect to their spatial distributions in the epithelium. The results were similar to those obtained with the OlfC probes since no co-expression of *ora* with taar genes was observed (Fig. VII-25D). Furthermore, *taar* genes are also expressed seemingly in a higher number of neurons when compared to *ora* genes and its distribution in the lamella ranges from very basal to high medial, i.e. clearly different from the more apical distribution of the *ora* expressed cells.

Figure VII-24 | ora genes show exclusive expression with respect to omp and s100z.
The OE expression patterns of ora versus the omp and s100z genes were compared using two-colour RNA in situ hybridization. OSNs labelled by a mixed ora3,4 probe were never labeled by omp-positive ciliated neurons (A,B). Similarly, the cells labeled by a mixed probe of the ora1,3 genes were never





positive for s100z (C-D). Arrows point to individual cells. Scale bars, 50 mm (A,C) and 10 um (B,D).

Figure VII-25 | ora genes show exclusive expression with respect to representatives of the OlfC and TAAR families.

The olfactory epithelium expression patterns of ora versus members of the OlfC and TAAR families were compared using two-colour RNA in situ hybridization. OSNs labelled by a mixed ora3,4 probe were never labeled by OlfCa1-positive neurons. (A,B). Similarly, the cells labeled by a mixed probe of the ora1,3,5 genes were never positive for OlfCj1 (C) neither for OlfCu1 (data not shown). Also, the expression pattern of a mixed ora1,3 probe shows exclusion from both the taar10- (data not shown) and taar12f-positive (D) OSNs. Arrows point to individual labeled cells. Scale bars, 50 mm (A-D).

7. *ora* genes co-express with Gαi and Gαo, consistent with signal transduction via these G-proteins

It is known that the mammalian branch of the Ora family, V1R receptors, couple with Gαi2 in rodents vomeronasal sensory neurons (VSN) (Berghard and Buck 1996; Berghard, Buck et al. 1996). More recently other studies in teleost fish revealed that crypt cells differentially express Gαq/11, Gαi or Gαo depending on the species, (Hansen, Rolen et al. 2003; Hansen, Anderson et al. 2004). Thus all these G-proteins appear to be good candidates to which the *ora* receptors may couple in their downstream transduction pathway. To test this hypothesis, I performed in-silico predictions using the PRED-COUPLE software (Sgourakis, Bagos et al. 2005; Sgourakis, Bagos et al. 2005), with all all the teleost fish Ora protein sequences as query sequences. The results are all summarized in Figure VII-26 and indicate that 68 or 79% of all teleost *ora* genes couple to either Gαo or Gαi (95 and 90% confidence level, respectively).

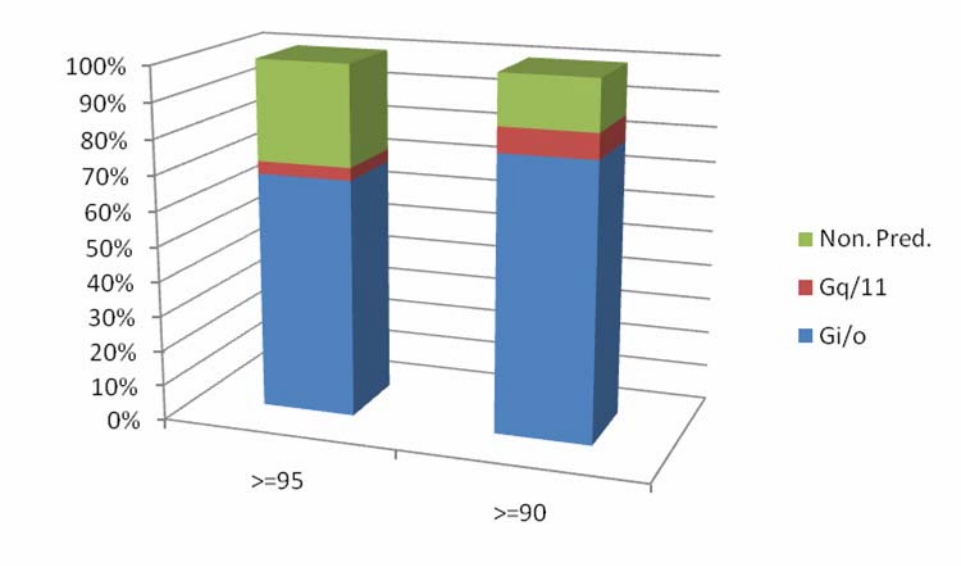


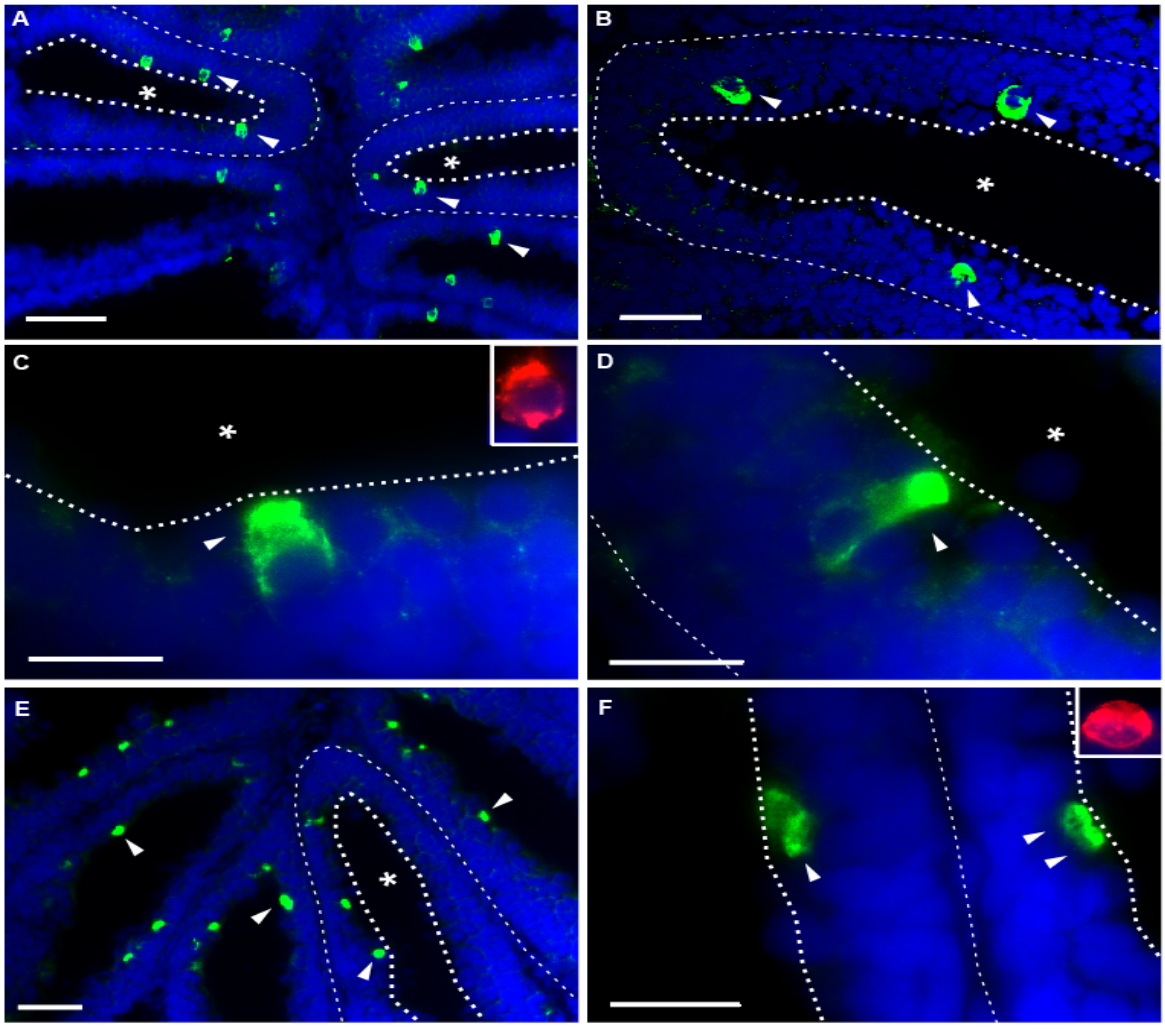
Figure VII-26 | ora genes are predicted to signal through Gαi/o.

In silico predictions show that most of the teleost ora genes might signal through Gαi/o. Predictions were obtained using the PRED-COUPLE2 software (***http://athina.biol.uoa.gr/bioinformatics/PRED-COUPLE2/***). The majority of the ora genes are predicted to signal through the Gαi/o subfamily. The method is based on a refined library of highly-discriminative Hidden Markov Models. Hits from individual profiles are combined by a feed-forward Artificial Neural Network to produce the final output. Sequences with prediction values with a threshold for significance lower than 90% were listed as "non-pred".

To examine this prediction, antisera directed against the two most highly predicted G-protein alfa-subunits, Gαi and Gαo, were used first to define their special distribution and second to examine whether the expression of these transduction related proteins overlaps with *ora*-positive OSNs labeled by fluorescent in-situ hybridization. Both Gαo and Gαi proteins were found to be expressed in sparse OSN distributed across the apical region of the lamella and seemingly scattered in a random fashion throughout the zebrafish olfactory epithelium (Figs. VII-27A,B, respectively), consistent with the crypt cell distribution as described above. Moreover, the Gαi staining seems to be present in the cytoplasm of cells with cryptic morphology (Figs. VII-27A-D), Surprisingly and in contrast to results described for catfish (Hansen et al 2003), Gαo is found to be present in a minute and extremely apical region of some cells, presumably the microvilli of crypt cells (Fig 29E,F).

Figure VII-27 | Gαi- and Gαo-immunoreactivity in the adult zebrafish olfactory epithelia.

Immunoreactivity (IR) of Go (A-D) and Gi (E,F) in the adult olfactory epithelium. Go is located in throughout the apical region of the lamellae (A,B). Gi shows a similar spatial distribution, confined also to the apical region of the lamellae. This distribution is similar to the one described for crypt cells. Moreover, both Go and Gi seem to label cells that are morphologically similar to crypt cells (C, E and F). The top tight corner panels in C) and F) represent S100-positive labeled cells and are included for matters of morphologic comparison. Arrows point to individual labeled cells. Scale bars, 50 um(A,E), 20 um(B) and 10 um(C,D,F).



The double-label experiments of *ora* RNA probes together with the antibody directed against the alpha subunits of either Gαi or Gαo showed that ora5 is expressed in the same cells as Gαo and that a mixed *ora1*,3 probe partially co-express with Gαi. This suggests that, like mammalian vomeronasal receptors, Oras may transduce signals by coupling to Gαo and Gαi, and thereby could lead to the activation of phospholipase C and the generation and elevation of intracellular IP3 levels (cf. Krieger and Breer 1999). However, further co-labeling experiments with more *ora* and G-proteins will be needed to establish whether the observed results for both *ora5*+Gαo and *ora1*,3+Gαi can be generalized to the whole *ora* family.

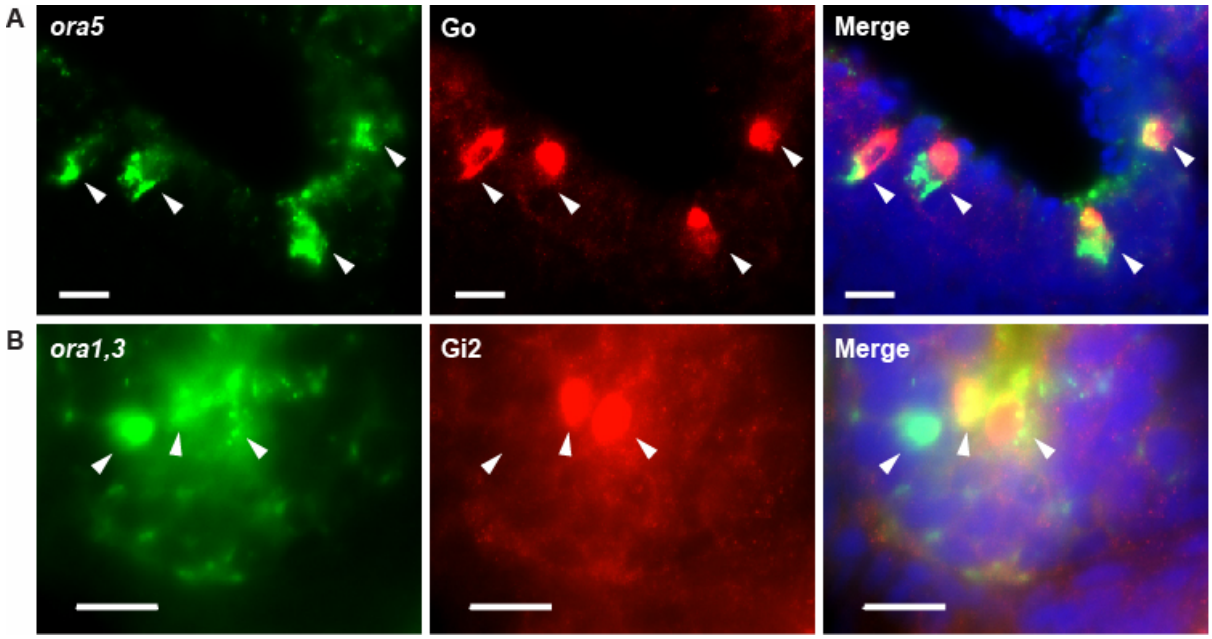


Figure VII-28 | *ora* genes co-express Gαo and Gαi.

Co-labeling experiments using ora probes and antibodies against Gαo and Gαi were performed. All the OSNs labeled by ora5 were also positive for Gαo (A). Similarly, a mixed ora1,3 probe showed co-expression with Gαi (B). However, not all the positive OSNs for Gαi seem to be labeled by the ora1,3 mixed probe and vice-versa. Arrows point to individual labeled cells. Scale bar 10 um.

VIII. DISCUSSION

A. Ora family

1. Ora genes constitute the fish homolog olfactory receptor gene family to the mammalian V1Rs

I have identified a novel olfactory receptor gene family in teleost fish, which I named *ora* for olfactory receptor genes related to class A GPCRs. The homologous mammalian V1R genes have not been formally assigned to any of the five major GPCR classes yet, but are most related to the class A or rhodopsin family of GPCRs (Schioth and Fredriksson 2005), this thesis). All mammalian V1R genes belong to one of three clades within the *ora* family. Previously a single V1R-related gene had been described in several teleost species (Pfister and Rodriguez 2005). The genes reported there correspond to *ora*1 (zebrafish) and *ora*2 (pufferfishes and medaka). The latest search for teleost V1R-related genes (Pfister, Randall et al. 2007; Shi and Zhang 2007) yielded further orthologs of *ora* 1 (medaka) and *ora*2 (zebrafish). *ora*3-6 presumably have been overlooked so far due to the low sequence identity between paralogs and their peculiar gene structure. The specific expression of all *ora* family members in the olfactory epithelium, and indeed in olfactory receptor neurons, as well as the relationship to mammalian V1R receptors support the identification of this novel family as olfactory receptors.

2. Ora genes display an extreme degree of negative selection among olfactory receptor gene families

Among olfactory receptor families the Ora family is unique for its small, rigidly maintained gene repertoire (no pseudogenes, no gene gains in five teleost species, only one gene loss in pufferfish), as well as the strict genomic arrangement in symmetrical gene pairs (for four of the six genes). Accordingly we found extensive negative selection in the *ora* genes, but no evidence for positive selection, in striking contrast to the situation in the mammalian V1R family (Grus, Shi et al. 2005) and also the OR and V2R families (Niimura and Nei 2005; Alioto and Ngai 2006). While the function of the *ora* genes is currently unknown, these features of the family suggest a small set of evolutionarily rather conserved ligands. If these ligands are pheromones as in the related mammalian V1R family (Boschat, Pelofi et al. 2002), they may be expected to be less species-specific than is usual for pheromones. Among known fish pheromones are some reproductive hormones, which are additionally recognized by the olfactory system (Friedrich and Korsching 1998; Stacey, Chojnacki et al.

2003). Such a double role engenders a double constraint, which could impede changes in the chemical structure of these pheromones during evolution. This hypothesis will be tested directly by analyzing the ligand binding properties of the novel *ora* gene family.

3. Evolutionary origin and dynamics of the Ora family in fish

The presence of all six *ora* genes in zebrafish an *Ostariophysi*, i.e. a rather primitive teleost) as well as in two more modern fish species (stickleback and medaka, both *Neoteleostei*) indicates their presence at least in *Otocephala*, while the existence of *ora*1, *ora*1-*ora*2 and *ora*3-*ora*4 orthologs in *Xenopus* suggests that the most recent common ancestor (MRCA) of tetrapods and teleosts already possessed the *ora*1-*ora*2 gene pair and at least one gene from the Ora3-Ora4 clade. This was confirmed by the presence of four Ora family members from jawless fish (lamprey) in all the 3 Ora clades: Pm OraW predates theOra1-Ora2 clade, Pm OraX predates the ORA3-Ora4 clade and Pm OraY and OraZ predate the Ora5-Ora6 clade. Furthermore, all these lamprey genes are not basal to the *ora* family, but cluster deep inside the phylogenetic tree. Additionally, the presence of orthologs for all the first *ora* genes (ora1 to ora4) in a cartilaginous fish implies that the Ora family expanded upon the segregation of jawless (*Agnatha*) from jawed fish (*Gnathostomata*). The fact that no orthologs for ora5 and/or ora6 were found in shark might suggest a loss of this gene(s) during evolution. However, this still remains to be elucidated due to the incomplete status of the elephant shark genomic database. Thus the origin of the family lies with very early vertebrate stages although individual family members only become reliably recognizable in the time period between the jawed/jawless segregation and the divergence of cartilaginous from bony fish around 560 to 530 million years ago.

The phylogenetic tree would support two ancient large scale genome duplications to generate the three clades observed here from a single ancestral gene. These duplications could correspond to the two whole genome duplications 590 and 440 million years ago (Meyer and Schartl 1999), i.e. before the teleost and tetrapod lineages split, or, alternatively, one of these in combination with a later whole genome duplication, which occurred in the ray-finned fish lineage after the teleost/tetrapod split, cf. (Hoegg, Brinkmann et al. 2004). Furthermore, the inverse orientation of the gene pair in the Ora1-Ora2 clade *vs.* the Ora3-Ora4 clade supports an independent origin of these two local duplication events. Indeed the genesis of the three clades preceded the genesis of the *ora* gene pairs (see above).

4. Frog Ora genes at a transition point between Ora and V1R evolution

In contrast to the fish Ora family, mammalian V1R family contain up to 150 genes (Rodriguez et al 2002; Zhang et al, 2004). In mouse, the large majority of the *V1r* genes are

organized in clusters, some containing over fifty *V1r* sequences with an average intergenic distance of 27Kb in a given cluster (Rodriguez, Del Punta et al. 2002). These clusters are located on different chromosomes (mainly 6,7,13 and 17, with a few on chromosomes 1,2 and X), with the largest *V1r* locus harboring 55 *V1r* genes, distributed along a 1,2Mb region on chromosome 13 (Rodriguez, Del Punta et al. 2002; Zhang, Rodriguez et al. 2004). In 2005, young et al described a similar number of V1r gene clusters in another rodent species, the rat (*Rattus norvegicus*), which are at syntenic genomic locations to those in mouse (Young, Kambere et al. 2005). However the same authors reported a different situation in dog and primate genomes (human and chimpanzee), where they find only a small number of V1Rs equivalent to the mouse V1R clusters and many other *V1r*-like sequences dispersed widely throughout the genomes (Young, Kambere et al. 2005). *Xenopus* constitutes thus the first example for the species-specific expansion of subfamilies frequently occurring in mammalian V1Rs (Grus, Shi et al. 2005). All but one of the 14 frog members in the Ora1-Ora2 clade, are located within a single genomic cluster. It is assumed that such expansions are caused by recurrent gene duplications. With this pattern of gene gains *Xenopus ora*s exhibit a feature common to all tetrapod olfactory receptor families and also to other teleost olfactory receptor families (Grus, Shi et al. 2005; Hashiguchi and Nishida 2005; Niimura and Nei 2005) (Shi and Zhang 2007). The striking absence of any gene gains in teleost *ora* genes is consistent with the notion of a radical shift both in the composition of the ligand repertoire and the physiological function of those ligands during the teleost-tetrapod transition.

Two other recent studies (Shi and Zhang 2007; Date-Ito, Ohara et al. 2008) confirmed the existence of a V1r-like family with a size of 21 functional genes in the frog genome. These frog V1r-like genes correspond to the fifteen Ora frog genes described in this thesis. The remaining six genes appear to be duplicate sequences of the same gene, with only some small variation in either the C- or N-terminal end of the molecule. In my experience this is caused by erroneous automated assignment of different transcripts for the same gene by the algorithms used, and is often cleared up in later, more refined versions of the genomic databases. Date-Ito and colleagues showed that all the frog V1R genes are expressed mainly in the middle cavity (MC) of the main olfactory epithelium, which is responsible for detecting water-soluble odorants (Date-Ito, Ohara et al. 2008). They also observed for two of the V1r frog genes a minor expression in the principal cavity (PC) of the main olfactory epithelium, which is responsible for detecting airborne odorants. In no case expression of the V1R genes was detected in the vomeronasal olfactory epithelium, where the rodent V1R genes are expressed. Taken together, all these results support both the assignment and naming of the frog V1R-like genes to the *ora* and not the V1R gene family (Date-Ito, Ohara et al. 2008)Furthermore, these data suggest that the change of the main expression site for

V1Rs from the MOE to the VNO might have occurred after the species divergence of mammalian vertebrates from amphibians.

5. Intron gains in the Ora gene family

The genomic arrangement of the *ora* genes presented two unexpected findings. In contrast to all mammalian V1R genes several instances of multiexonic organization are found for the *ora* genes. However, the ancestral genomic structure seems to be monoexonic for all *ora* genes. This is supported by the unequivocal presence of three and two intronless *ora* genes in both lamprey and shark, respectively. The intronless gene structure is maintained in the tetrapod relatives of the Ora1-Ora2 clade, the amphibian and mammalian V1R genes (Saito, Mimmack et al. 1998; Rodriguez, Feinstein et al. 1999; Del Punta, Rothman et al. 2000).

In stark contrast, *ora*4 possesses two exons and for *ora*3 four exons are predicted in all five fish species examined. Also the intron gain for the multiexonic *ora* genes seem to have occurred upon segregation of the cartilaginous and bony fish lineages. However, for ora4 the hypothesis that the intron gain occurred already in the cartilaginous lineage cannot be excluded and can only be evaluated upon the release of a more complete version of the shark database. The strict conservation of intron-exon structure may indicate the existence of regulatory elements within these introns. The size distribution of the exons is rather conserved between species and intron/exon borders are accurately maintained.

Interestingly, none of these four intron/exon borders are found in the monoexonic *Xenopus* Ora3-Ora4 clade member (Xt *ora*15, data not shown) nor in the intronless Ora3-Ora4 clade member found in *Petromyzon marinus*. The most parsimonious explanation for the absence, presence, and location of the different intron/exon borders in *ora*3 *vs. ora*4 is an independent gain of introns posterior to the genesis of the *ora*3-*ora*4 gene pair, but prior to the teleost speciation taken into account here.

Several publications recently have demonstrated a reduction in the percentage of intron-containing genes in higher vertebrate GPCRs (Bryson-Richardson, Logan et al. 2004). This is supported by the evidence presented here. In teleost fish about one half of all *ora* genes contain introns, whereas in *Xenopus* no introns are present (data not shown), and in mammalian V1Rs no incidence of introns has been reported. However, the explanations given for this reduction appear only partially applicable to the *ora*/V1R group of genes. A loss of introns has been suggested by (Bryson-Richardson, Logan et al. 2004), whereas the genesis of new, intronless genes in higher vertebrates has been emphasized by (Fridmanis, Fredriksson et al. 2007).The large expansion of the intronless Ora1-Ora2 clade in *Xenopus* is consistent with the interpretation by (Fridmanis, Fredriksson et al. 2007). We do not

observe any intron loss in the *Ora*/V1R family; on the contrary my data support the gain of four introns early in the teleost lineage. These intron gains in the slowly evolving teleost Ora family support the observation by (Carmel, Rogozin et al. 2007; Carmel, Wolf et al. 2007), who found that such gains preferentially occur in evolutionarily conserved genes.

6. Genomic linkage of ora-gene pairs

The other genomic feature novel for olfactory receptor genes is the occurrence of tightly linked symmetrical *ora* gene pairs, which are conserved throughout teleost evolution. The emergence of these gene pairs is not completely resolved, but the deduced whole genome duplications in the vertebrate lineage cannot be responsible, as the gene pairs are a local structure. In other cases of such gene pairs regulatory elements of one gene have been shown to lie in the other gene, enforcing linked evolution. Quite possibly some of these elements might even be shared among both genes of the pair (cf. (Sumiyama, Irvine et al. 2002), which would explain the maintenance of the symmetrical arrangement of the gene pairs (head-to-head and tail-to-tail for ora1-ora2 and ora3-ora4, respectively). It is conceivable that initially *ora*5 and 6 also occurred as such a gene pair, but that their association was degraded long before teleost speciation. Mutual dependency of expression for the gene pair would result in a slowed down evolution of the Ora1-Ora2 and Ora3-Ora4 clades compared to the Ora5-Ora6 clade, whose ortholog identities indeed tend to be distinctly lower than that of the other four genes. Conversely, the rapid gene expansion in the *Xenopus* Ora1-Ora2 clade predicts a loss of the gene pair arrangement in this species. Indeed, no pairwise arrangement is observed in this species. Even though thirteen of the fifteen *Xenopus ora* genes, including the *ora*1 ortholog, are clustered together in a 272 kb small genomic region, the *ora* genes nearest to *ora*1 are 35 and 45 kb apart, and all *ora* genes except the most distant one share the same orientation. Interestingly, for many genes in this cluster their genomic location correlates with their phylogenetic relationship (Fig. VII-1), consistent with a recurrent duplication of the cluster element most distant from *ora*1.

B. TAAR family

1. Origin and delineation of the TAAR gene family from the classical aminergic receptors

TAAR family is the most recently described olfactory receptor gene family in mammalian species (Liberles and Buck 2006), and unlike the other three families of olfactory receptor genes (OR, V1R, V2R) have not undergone major radiation in mammals. I was thus interested in defining the characteristic properties of the family responsible for the rapid evolution observed in teleosts. Currently rather completely sequenced genomes are available for several teleost species and I have taken advantage of this large improvement in databank quality to help establishing the complete *taar* gene repertoire in five teleost fish species.

Previous estimates of family size have been either too low, *cf*. (Gloriam, Bjarnadottir et al. 2005), presumably due to incomplete databases or too high due to inadequate delineation of the *taar* gene family from the related aminergic receptors, *cf*. (Hashiguchi and Nishida 2007). In my experience it is necessary to include representatives from all major aminergic receptor families to obtain a proper delineation of the *taar* gene family which is supported by the presence of the characteristic TAAR fingerprint motif (Lindemann and Hoener 2005). In this analysis all lamprey receptors previously considered TAARs (Hashiguchi and Nishida 2007) clearly segregate with teleost and tetrapod aminergic receptors, not with teleost nor tetrapod *taar* genes. Consequently, the origin of the TAAR family is more recent than previously thought. The discovery of shark *taar* genes places the origin within the MRCA of cartilaginous and bony fish. This fact makes the TAAR family more recent than the Ora family and probably the youngest one among all the described olfactory receptor gene families.

2. Massive radiation of an olfactory receptor gene family only in teleost fish

All the other three olfactory receptor gene families (Ora/V1R, OlfC/V2R and OR) are characterized by a huge expansions upon transition from water to land. The outcome is the existence of massive radiations of these families in both frog and mammals compared to fish species. However the scenario in the TAAR family proved to be very different. The TAAR family appears to have remained rather small for long periods of evolution, with fairly recent bouts of extensive local gene duplications which left their traces in the strong correlation between genomic and phylogenetic neighborhood between genes. The expansion(s) achieved massive proportions during the expansion of the teleost fish class III genes. This is

consistent with the teleost-specific expansions after the segregation of the fish and tetrapod lineages in other teleost olfactory receptor gene families (Niimura and Nei 2003; Hashiguchi and Nishida 2005; Niimura and Nei 2005; Alioto and Ngai 2006; Hashiguchi and Nishida 2006; Niimura and Nei 2006; Niimura and Nei 2007). Surprisingly, nearly no expansion of the family was observed in any of the mammalian families analyzed, in stark contrast to the situation witnessed in other olfactory receptor gene families. In addition, some gene losses shaped the *taar* gene repertoire, most notably the loss of TAAR1 in all neoteleosts examined here (Hussain et al). The rapid evolution in the *taar* gene family is paralleled by intron dynamics unprecedented for olfactory receptors (*cf*. (Niimura and Nei 2005). Two independent intron gains and two independent intron losses, all exclusively in the neoteleost *taar* genes of class III underscore the fast evolution especially in this class (Hussain et al).

3. Unprecedented level of positive selection in an olfactory receptor gene family

The major radiation in class III goes hand in hand with weak to nearly absent global negative selection. However, as expected from recent and massive gene radiations, three subfamilies of class III, all of them species-specific, show extended positive selection at individual sites, well beyond that observed in ORs and V2R-like OlfC genes (one to two sites (Alioto and Ngai 2005; Alioto and Ngai 2006), determined using the same algorithm. The other olfactory receptor gene family present in fish (Ora family) constitutes an extreme case of negative selection and did not show any evidence of positive selection in its sequences. In contrast, the TAAR genes showed up to 20 positively selected sites per gene, thus constituting an extreme case of positive selection among all the olfactory receptor gene families. Positive selection is usually taken as evidence that divergence in sequence is selected for. For olfactory receptor genes (ORs) this has been argued as a mechanism to maximize the odor space recognizable by the receptor repertoire. The presence of extensive positive selection in the teleost *taar* gene family is consistent with a role as olfactory receptor genes.

4. Taars as olfactory receptor genes

This assumption that TAARs are olfactory receptor genes is supported by the expression of all but one *taar* genes studied in sparse olfactory receptor neurons. The intermediate position of labeled neurons in the lamella is consistent with an expression in ciliated receptor neurons (cf (Sato, Miyasaka et al. 2005), which would be analogous to the mammalian situation. The ligands of teleost class I and II TAARs may include various amines (*cf*. mammalian TAARs (Lindemann and Hoener 2005; Liberles and Buck 2006), consistent with

the detection of amines by the fish olfactory system (Rolen, Sorensen et al. 2003). It is intriguing that TAAR1, the most ancient *taar* gene, is not an olfactory receptor - it appears that the TAAR family has begun its existence with a function different from the one currently emphasized.

5. Taars and ora gene families constitute two opposite ends of the "olfactory gene repertoire spectrum"

Ora genes are evolutionary old, with representatives already in jawless fish, evolve slowly and exhibit strong negative selection. In contrast, *taar* genes first appear in jawed fish, evolve rapidly (especially in teleost fish), and show extended positive selection. Moreover, *ora* genes display a rigidly maintained genomic arrangement, and no genesis of new genes during all of the jawed fish evolution, in contrast to both the tetrapod V1R genes and the *taar* gene repertoire, which is extremely species specific in teleosts (if not in mammals).

It will be interesting to identify the ligands for both families of receptors to investigate their respective physiological roles which might be distinctively different.

C. Cellular and molecular properties of crypt cells and ora receptors

In this section I will discuss the identification of the particular type of receptor neurons which express the *ora* genes, the expression pattern of *ora* genes at the cellular level, and initial results obtained in the identification of the signal transduction pathway for Ora receptors.

1. The s100 gene family might contain a crypt cell molecular marker

Crypt cells constitute one of three subtypes of olfactory receptor neurons (the other two being microvillous and ciliated cells). Previously it has only been possible to identify crypt neurons by morphology and cross-reactivity of an antibody raised against mammalian S100 protein. Functional and ontogenetic analysis of crypt neurons has correspondingly been hampered. I have identified a potential molecular marker for crypt neurons among the S100 family, S100Z, however, due to the duality of results obtained with different methods, the assumption that S100Z is co-expressed with S100 antibody staining currently can not be verified unambiguously. Conceivably, the immune staining of the S100 antibody could be caused by a different *s100* gene. Of nine other *s100* genes analysed so far in the three day old larvae, seven *s100* genes can be safely excluded as causative for the observed S100 staining in the placode, since their expression pattern does not include the olfactory placode. One gene, *s100i*, is expressed in the placode, but also in skin Fig. 20, *cf*. (Kraemer, Saraiva et al. 2008). However, it does not appear to cause the S100 antibody labeling, since

preliminary experiments showed a broad expression pattern in the adult olfactory epithelium from very apical/dorsal to basal, and also not limited to the sensory area of the neuroepithelium (Oka, Y, personal communication), unlike the antibody staining. It is possible that one or more of the remaining four genes that were not yet analyzed shows an expression in the OE. For time constraints this analysis could not be completed so far. If successful, this finding will complete the set of molecular markers for the three different olfactory receptor neuron subtypes, with TRPC2, an ion channel, useful as marker for microvillous receptor neurons (Sato, Miyasaka et al. 2005) and olfactory marker protein (Celik, Fuss et al. 2002) for ciliated neurons.

2. An hypothesis about the functional role of s100 genes in the olfactory epithelium

Ca2+ is known to play an instrumental role in olfactory signal transduction (Menini 1999) (Matthews and Reisert 2003), neuronal stress response and xenobiotic metabolism (Moncada and Bolanos 2006), as well as in apoptosis and regeneration (Schwob 2002) (Beites, Kawauchi et al. 2005). In many cases, Ca2+ is not interacting directly with its molecular target, but its effects are mediated by Ca2+-binding proteins (Bastianelli and Pochet 1995; Bastianelli and Pochet 1995; Bastianelli, Polans et al. 1995; Bastianelli, Takamatsu et al. 1995; Miwa, Kobayashi et al. 1998; Mammen, Simpson et al. 2004). In particular, calmodulin is highly expressed in ORNs and has been shown to regulate all key steps of the transduction cascade located in the cilia (Anholt and Rivers 1990; Biffo, Goren et al. 1991).

In contrast, the S100 family of calcium-binding proteins is much more cell- and tissue specific. S100 proteins take part in many cellular processes which may be divided into five major groups: a) modulation of the activity of some protein kinases b) modulation of other enzymatic activities c) maintenance of cell shape and motility d) modulation of signal transduction pathways and e) regulation of calcium homoeostasis (*cf.* (Santamaria-Kisiel, Rintala-Dempsey et al. 2006). S100 genes code for small, cytoplasmic or secreted proteins that exhibit a two domain structure each consisting of a calcium-binding loop flanked by two alpha-helices (Marenholz, Heizmann et al. 2004). The N-terminal domain is a low affinity S100-specific domain, whereas the other one is a classical EF-Hand. A variable hinge region connects both domains. Structural examination revealed a hydrophobic region being exposed upon calcium binding which is thought to interact with hydrophobic regions of the target proteins. The calcium-binding domains also exhibit affinity for other divalent cations such as copper and zinc (Heizmann and Cox 1998).

The function of S100 in crypt neurons is not known, but it is interesting to note that one member of the S100 family, S100B, is involved in signal transduction from a membrane-bound guanylate cyclase receptor in the visual and the gustatory system (Sitaramayya 2002; Duda and Sharma 2004). Consistent with this, one or more members of the S100 family might be involved directly in signal transduction from the olfactory receptors of the crypt cells (see below). Alternatively S100 genes could serve to modulate olfactory pathways, e.g. olfactory imprinting (Harden, Newton et al. 2006) or cellular differentiation processes.

3. Deorphanization of crypt neurons and their putative functional role

To date no olfactory receptors of crypt cells have been identified. I showed that all the members of the *ora* gene family (Saraiva and Korsching 2007) are expressed in crypt cells, thus deorphanizing the third type of OSNs present in the fish epithelium. *Ora*-expressing cells are always a subpopulation of S100 antibody positive neurons, as expected. Whether all S100-labeled cells express an *ora* gene is not clear so far, and would require further quantitative analysis to be determined. Moreover, I also found that the vertical distribution of *ora*-expressing cells in the adult olfactory epithelium resembles the vertical distribution of the crypt cells (data not shown).

The task of crypt cells themselves is not clearly established at the moment. They have been shown to respond to (rather high concentrations of) amino acids (Schmachtenberg 2006), but other studies implicate them in the detection of sexual pheromones (Weltzien, Hoglund et al. 2003; Hamdani el and Doving 2006; Lastein, Hamdani el et al. 2006). It is to be expected that the availability of a molecular marker will accelerate the progress considerably and allow studying the development, differentiation and mature function of crypt neurons.

As a first result I report the presence of crypt cells already in embryonic stages of zebrafish development. Previously these cells had only been described in juveniles and adults from several species of bony fish (Zeiske, Kasumyan et al. 2003; Yamamoto, Mori et al. 2004), but recently crypt cells were also found in the cartilaginous fishes *Scyliorhinus canicula* (Ferrando, Bottaroa et al. 2007) and *Raja clavata (*first described 15 weeks post fertilization, well before hatching (Ferrando, Bottaroa et al. 2007). Furthermore, the developmental onset of the *ora* genes seems to be close to that of the crypt cells, since ISH of three *ora* genes performed in 5dpf zebrafish larvae revealed an expression in the epithelium. The presence of both the crypt cells and expression of the *ora* genes already in late embryonic and early larval stages is consistent with an additional function of crypt cells beyond the suggested detection of reproductive pheromones. At this developmental stage, shortly after hatching, young fish start to explore their environment and begin with foraging.

Crypt cells might play a role in mediating or modulating those behaviors, thus making amino acids good ligand candidates for the *Ora* gene family.

4. Monogenic expression of ora genes

Monogenic expression of odorant and vomeronasal receptor genes, the one receptor neuron - one receptor gene rule, is thought to be the standard operating procedure in the olfactory system. Nevertheless, a systematic analysis of this dogma is still missing. Previous data suggest a possible co-expression of different V1rs in single sensory neurons, based on the fact that probes recognizing about 35% of the complete V1r repertoire (families V1ra, V1Rb and V1Rd react with most of the Gai2-expressing VSNs, (Pantages and Dulac, 2000). Similarly, another study reported that in a subpopulation of rodent VSNs, members of a distinct receptor subfamily (V2R2) are broadly co-expressed in the same cells as other V2Rs (martini et al 2001). Interestingly, the V2R2 is the rodent ortholog of the fish OlfCa1 gene used in my study as a general marker for microvillous neurons. This raises the possibility that the co-expression of several members of the ORA family in the same OSNs might also be possible. Furthermore, the peculiar genomic arrangement of the two fish *ora* gene pairs might reflect a possible common regulation of both genes, which could result in the co-expression of ora1/ora2 or of ora3/ora4 in the same OSN. However, all the pairwise combinations I tested among ora genes did not show co-expression of two genes in the same OSN. This is consistent with the monogenic rule of expression also being valid for *ora* genes (refs). The monogenic rule was in fact recently confirmed for the mammalian V1R orthologs of the Ora family. In that study, Ropollo and colleagues (embo Rodriguez 2007) showed that none of the V1Rs from the other subfamilies ever co-expressed with V1rb2, which might have been expected from the results reported by Pantages & Dulac (2000) described above.

Why would one OSN only express one olfactory receptor? The co-expression of several olfactory receptor genes, even if highly related, may be incompatible with the formation of functional odotopic maps in the olfactory bulb. Furthermore, it was shown that in the rodent main olfactory system, minor changes in olfactory receptor sequence lead to different axonal responses to guidance signals in the olfactory bulb (Feinstein and Mombaerts 2004). It was therefore suggested that a dual pressure may act on the mechanisms leading to the expression of a single V1r ‘morph’ per VSN: on one side selective pressure to avoid a complex decoding challenge, and on the other side, to facilitate axonal wiring processes (Roppolo, Vollery et al. 2007). In an additional set of observations it has been shown that co-transcription of a pseudo and a functional olfactory receptor gene can take place in the same OSN, while the coexpression of more than one functional olfactory receptor gene is found

only rarely (Rawson and LaMantia 2007). Thus, it is hypothesized that a mechanism based on negative feedback mediated by the expressed olfactory receptor, and/or negative selection against OSN expressing more than one functional olfactory receptor might be the responsible (Serizawa, Miyamichi et al. 2003; Feinstein, Bozza et al. 2004; Feinstein and Mombaerts 2004; Lewcock and Reed 2004; Shykind, Rohani et al. 2004).Such regulatory mechanisms, where the receptor itself is the key regulator of the process that ends in monogenic expression, are present in both the mammalian main olfactory and vomeronasal systems. Holding true, this could help to explain why, despite possibly sharing regulatory elements, *ora* genes belonging to the same gene pair are never co-expressed in the same OSNs.

5. G-proteins and cell type

In rodents, ciliated cells of the main epithelium express OR that couple to Gαolf. In contrast the microvillous cells of the vomeronasal organ express V1R and V2R families of receptors that couple to Gαo and Gαi, respectively (Dulac and Axel 1995; Dulac 2000) (Matsunami and Buck 1997). Likewise, in fish species the ciliated cells express ORs coupled to Gαolf, whereas microvillous ORNs express V2R-family receptors along with either Gαi, Gαo or Gαq (Hansen, Rolen et al. 2003; Hansen, Anderson et al. 2004). Furthermore, crypt cells, depending on the fish species, co-express either Gαo alone (Hansen, Rolen et al. 2003) or co-express Gαq and Gαo in their microvilli and cilia, respectively (Hansen, Anderson et al. 2004). Even though in fish, unlike mammals, these three types of OSN co-exist in one single olfactory organ (the OE), their projections segregate in the olfactory bulb, consistent with the concept of odotopic representation in the OB (Doving and Selset 1980; Friedrich and Korsching 1997; Nikonov and Caprio 2001).

My results show that the antibody against Gαo labels the somata of cells located in the apical region of the lamella, that anatomically resemble crypt cells. This is consistent with the results reported for the channel catfish, which expresses Gαo in the cytoplasmic region of the crypt cells (Hansen, Rolen et al. 2003). The antibody against Gαi labels possibly the same apical population of cells, but surprisingly Gαi is restricted to the very apical region (maybe microvilli) of these cells. Both the numbers and spatial distribution of cells labeled with Gαo and Gαi resemble those of crypt cells, although a putative labeling of the microvillous cells cannot be excluded with the current data. However, crypt cells of the closely related goldfish co-express Gαq in the microvilli and Gαo in the cilia (Hansen, Anderson et al. 2004). So far it is not clear whether species differences account for the observed differences in G proteins expressed. However, the co-expression of two G proteins

has now been consistently found in several species, and may indicate the existence of two different transduction pathways in teleost crypt cells.

6. G-proteins and Ora receptors

Consistent with previous observations in several fish species I found that the spatial distribution in the vertical dimension of the OE (dorsal to ventral, sometimes called apical to basal) is different for the three types of OSNs. Ciliated cells show the broader distribution, ranging from very dorsal to very ventral (data not shown). In contrast, crypt cells show the narrowest spatial distribution pattern, concentrated dorsally in the vertical dimension and near the midline raphe of the epithelium in the horizontal dimension ((Hansen, Rolen et al. 2003; Hansen, Anderson et al. 2004), and data not shown). Microvillous cells exhibit an intermediate and more variable pattern in that they tend to lie mainly dorsally within a lamella, but subsets may extend more ventrally (cf. (Morita and Finger 1998; Eriksson 1984).

Preliminary co-label experiments of several *ora* genes with either Gαo or Gαi showed co-expression in both cases. While all ora5-expressing cells appeared to be a subset of Gαo-positive cells, not all the ora1,3-expressing cells were positive for Gαi. Currently, technical problems cannot be excluded as explanation for the latter. Also, so far it is not clear, whether all the Gαi and Gαo-positive cells co-express *ora* genes, and it cannot be excluded that Gαi may also label another cell type in the fish olfactory epithelium, e.g. microvillous cells. Further co-labeling experiments of G proteins with the marker genes for crypt, ciliated and microvillous cells will be required to solve these questions. In any case, the preliminary data are consistent with the hypothesis that both G-proteins are present in, possibly the same, crypt cell neurons. This raises another question: Why would crypt cells have two different transduction signaling pathways? The most straightforward answer is that these OSNs may express more than one olfactory receptor. Since I could exclude co-expression among *ora* genes in the same OSN, the second type of receptors would possibly belong to a different family. In any case, the hypothesized existence of two distinct odor signal transduction pathways in one crypt cell would be consistent with the unusual morphology of crypt cells, which bear two types of sensory appendages, cilia and microvilli.

D. Conclusion and outlook

Taken together I have identified a novel family of six olfactory receptor genes in fish (Ora family) and contributed to building and characterizing the total repertoire of another one (TAAR). High conservation of both sequence and genomic arrangement, as well as strong negative selection characterize the *ora* genes, and set them apart from both their mammalian relatives and the teleost *taar* genes, which show extended positive selection.

Albeit a thorough understanding of the role of these two receptor gene families will require among others the identification of their ligands it appears reasonable to assume that their physiological roles will turn out to be very different, in line with their opposing phylogenomic characteristics.

By showing the *ora* gene expression in crypt cells I have de-orphanized this peculiar type of olfactory receptor neurons. My findings are consistent with the hypothesis that different populations of OSNs, utilizing different receptors and transduction cascades, arose early in vertebrate evolution. For a long time, these different types of ORNs remained co-localized in a single olfactory organ. However, the heterogeneous distribution of these OSN populations across the neuroepithelium may amount to the initial stage of segregation of different OSN populations into distinct epithelial chambers, such as the VNO and MOE in tetrapods.

IX. III. MATERIAL AND METHODS

A. Biological Materials

1. Animals

In this study wild-type zebrafish of the Ab/Tü strain (mix between the Oregon and Tübingen strains). Adult zebrafish (Danio rerio) were kept in group tanks at a day/night rhythm of 14/10 h at a water temperature of 28°C and fed daily dry flake foods and brine shrimp (artemia; Brustmann, Oestrich-Winkel). The aquaria were filled with a one-to-one mixture of demineralized water and tap water. Zebrafish embryos and larvae were kept in petridishes at a density of about 50 embryos/petridish in embryo medium (E3: 5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl2, 0.33 MgSO4, Methylenblue 5-10%) at 28°C without feeding for the first five days post fertilization (dpf). Afterwards, they were raised in 2 l containers. Until an age of two weeks they were fed a special food for fish larvae (TetraMin Mini, TETRA), and afterwards with artemia. In order to elicit controlled reproductive activity, selected females and males were put into a separate tank. Early in the following morning, fertilized eggs were collected. The embryos were then raised and collected at 24h intervals for histological and immunohistochemical processing. Embryos fixed at a stage older than 24 h postfertilization (hpf) were raised in 2 mM 1-phenyl-2-thiourea (PTU) in embryo medium after the epiboly stage (about 12 h) to prevent pigmentation. The embryonic and larval stages used for all investigations reported here ranged between one and 21 days postfertilization (dpf).

2. Bacterial Strain

For DNA amplification the *Escherichia coli* XL1 Blue MRF' (Stratagene, Heidelberg) bacterial strain was used.

B. Chemicals and Supplies

Chemicals used in this study were from Ambion (Austin, USA), Amersham Pharmacia Biotech (Freiburg), Applichem (Darmstadt), JTBaker supplied by Fisher Scientific (Schwerte), Biozym (Hessisch Oldendorf), Calbiochem (Darmstadt), Difco (Detroit, USA), Fluka (Neu-Ulm), Merck (Darmstadt), Molecular Probes (Leiden, NL), Roth (Karlsruhe), Serva (Heidelberg), or Sigma (Deisenhofen) unless stated otherwise in the text.

1. Enzymes

Restriction enzymes used were either from Amersham Pharmacia Biotech (Freiburg) or New England Biolabs (Schwalbach, Taunus). T4 DNA Polymerase, Taq DNA Polymerase, Expand High Fidelity Taq Polymerase, Expand Long Template Taq Polymerase, T4 DNA ligase, T3, T7, and SP6 RNA Polymerase and the Klenow enzymes were purchased from Roche Biochemicals (Mannheim). Reverse Transcriptase Superscript II was from Invitrogen Life Technologies (Karlsruhe) or from Bioline (Luckenwalde). RNase-free DNase RQ1 was from Promega (Mannheim), RNaseA and Proteinase K were purchased from Sigma or Roche Biochemicals (Mannheim).

2. Nucleotides

Nucleotides for PCR, reverse transcription, and in vitro transcription were purchased from Invitrogen Life Technologies (Karlsruhe) or Operon (Cologne).

3. Plasticware

All disposable plasticware like 15 ml and 50 ml Falcon tubes, 6-, 24-, 48-, 96-well plates, petridishes in various sizes were from BD or Castor, purchased from Fisher Scientific or BD biosciences. 96-well plates Polyfiltronics for colony PCR were from Whatman (supplied by Fisher Scientific), 0.2 ml PCR tubes and sterile pipette tips were from M_P supplied by Fisher Scientific. Non-sterile pipette tips were supplied by LaFontaine (Forst/Bruchsal) and Labomedic (Bonn).

4. Preparation of Solutions

Solutions were prepared with didiestilated water. Solutions were autoclaved for 20 min at 121 bar or filter sterilized (0.2-0.45 µm pore diameter). Glassware was autoclaved and oven baked for 2 h at 180°C. For RNA-work, solutions and water were treated with 0.1% diethylpyrocarbonate (DEPC), shaked vigorously and mixed for about 20 min on a magnetic stirrer to bring the DEPC into solution. The solutions were then autoclaved to remove any trace of DEPC. Tris buffers cannot be treated with DEPC because it reacts with primary amines. DEPC decomposes rapidly into CO2 and ethanol in the presence of Tris buffers. Therefore, Tris buffers were prepared by using water that has been treated with DEPC first. Most of the standard stock solutions like EDTA, Tris, TAE, TBE, TE, PBS, SDS, SSC, NaOAc, and culture media like LB and SOC were prepared as described in (Sambrook J 1989). All solutions used are named in the text.

C. Plasmids and Vectors/Properties

The plasmids used were the following:

- pGEM-T, Promega 3 kb; B/W; T vector; ampicillin resistance
- pBluescript II KS(+) ,Stratagene, 2.96 kb; B/W; ampicillin resistance
- pDrive , Qiagen; 3.85 kb; B/W; ampicillin and kanamycin resistance

B/W: blue/white selection possibleDescription

D. Oligonucleotide Primers

Oligonucleotide primers were purchased from Operon (Cologne) or Invitrogen Life Technologies. The primers were delivered or dissolved at a standard concentration of 100 mM. Working dilutions were prepared at a concentration of 10 mM and stored at -20°C. Primers were used for different purposes like sequencing, cloning, and for preparation of probes, by addition of T3-RNA Polymerase binding site. All used primers are listed below:ence (5'_3') T (°C) Supplier Purpose

- M13-Fwd, GTAAAACGACGGCCAGT,
- M13-Rev, AACAGCTATGACCATG,
- T3, TATTAACCCTCACTAAAGGGAA,
- T7, GTAATACGACTCACTATAG,
- SP6, CATTTAGGTGACACTATAG,
- Dr *actin*-Fwd, CCCCATTGAGCACGGTATT,
- Dr *actin*-Rev, AGCGGTTCCCATCTCCTG,
- Dr *ora*1-Fwd, ATGGACCTGTGTGTCACCATCAAAGGCGT,
- Dr *ora*1-Rev, TCATGGAAGTCCACATGGCAGAAG,
- Dr *ora*2-Fwd, ATGATTGCGGAGGCTGTG,
- Dr *ora*2-Rev, TCCACGTTGATGGCGTTC,
- Dr *ora*3-Fwd, ATGGCGCCTCAAAAGAAACCC,
- Dr *ora*3-Rev, AGATGAAGGCAGGGATGGAGT,
- Dr *ora*4-Fwd, ATGTCTGAGGTCCTGACGGTG,
- Dr *ora*4-Rev, GTGGTGCAGCTAATCACCATC,
- Dr *ora*5-Fwd , ATGCAGCTCCAAGACTGGGTT,
- Dr *ora*5-Rev, GGAGTTGGGAATTTTTCCTCA,
- Dr *ora*6-Fwd, ATGGTGATGGAGCAGATACAGGTGAATC,
- Dr *ora*6-Rev, AGCACACTCGTCACCGTGA,
- Dr *OlfCa1*-Fwd, GGGCCTTTTGAGAACGACACATG,
- Dr *OlfCa1*-Rev, CAGATTTGCCCATTAGCGAAGAGAG,
- Dr *OlfCj1*-Fwd, TGAGGGTTGGATCACGTACA,

- Dr *OlfCj1*-Rev, ATGCGTCATACAAGCCAATG,
- Dr *OlfCu1*-Fwd, GCTCCTGGTTGAAGTTGCTC,
- Dr *OlfCu1*-Rev, ACAGGCTCTCCATTGGTGTC,
- Dr *taar1f*-Fwd, ATGGATCTCTGTTATGAGGCG,
- Dr taar1f-Rev, GATGTAGAAGGAAAACACAGAGGTG,
- Dr *taar10f*-Fwd, ATGGACCTAAGCAATTCA,
- Dr *taar10f*-Rev, TACCATCGCAAATCCAACAA,
- Dr *taar12f*-Fwd, ATGAAGCCTTCAAATGAGAC,
- Dr *taar12f*-Rev, GTCACAAATGGCCCAGTACC,
- Dr *taar13c*-Fwd, ATGGATTTATCATCACAAG,
- Dr *taar13c*-Rev reverse, AACTGACCACAAGGCATTGAA,
- Dr *s100z*-Fwd, TAAACTGGAGGGAGCAATGG,
- Dr *s100z*-Rev, TCCAGCACTCAGTTTACGAT,
- Dr *omp*-Fwd, CAAGGACACACAGTAGACGC,
- Dr *omp*-Rev, GGAACAGACTGACCAGAAGAG.

The templates for the RNA probes were amplified from the cloned DNA using the same forward primers as above and reverse primers with the described T3 promoter site (TATTAACCCTCACTAAAGGGAA) attached to their 5' end.

E. Primary Antibodies

- S-100 polyclonal rabbit Dako, 1:200
- $G\alpha_{i-3}$ polyclonal rabbit IgG, Santa Cruz Biotechnology, 1:500
- $G\alpha_{o}$ polyclonal rabbit IgG, Santa Cruz Biotechnology, 1:500
- Anti-DIG sheep Fab fragment coupled with alkaline phospatase, Roche, 1:500-1000
- Anti-Flu sheep Fab fragment coupled with alkaline phospatase, Roche, 1:500-1000

F. Secondary Antibodies

- Donkey α-rabbit, Alexa Fluor 488 coupled, Molecular Probes, 1:200
- Donkey α-rabbit, Alexa Fluor 594 coupled, Molecular Probes, 1:200

G. Dyes, Substrates, Embedding Media and Counterstains

1. Alkaline Phosphatase Substrates

- NBT/BCIP (Roche Biochemicals) blue/violet chromogenic precipitate
- HNPP/Fast Red (Roche Biochemicals) red chromogenic and fluorescent precipitate

2. Horseradish Peroxidase Substrates

- Diaminobenzidine (DAB) (Roche Biochemicals) brown chromogenic precipitate
- Alexa Fluor 488 and 594 tyramide from the TSA kit with HRP-Streptavidine (Molecular Probes, Invitrogen detection technologies)

3. Embedding Media

- Vectamount (Vector) embedding medium for chromogenic substrates
- Vectashield (Vector) embedding medium for fluorescent substrates and dyes; good bleaching retardant

4. Dyes and Counterstains

- Vectashield contains DAPI that is used as counterstaining for the nuclei.

H. Equipment

Unless stated otherwise, general lab equipment was usedfor the molecular and cell biology techniques, including – balances, centrifuges, electrophoresis equipment, electroporation pulser, heating blocks and plates, hybridization and incubation ovens, micropipettes, PCR and gradient thermocyclers, pH meter, shakers, sterile hood, UV-transilluminator, vortexes and waterbaths. Fresh frozen sections were obtained using the Cryostat CM 1900, Leica. A Nikon SMZ-U binocular microscope equipped with Nikon CoolPix 950 digital camera attached was used to document whole mount images. A Zeiss AxioVert microscope with an attached Diagnostic Instruments Spot-RT camera was used to document non-fluorescent images. A fluorescent microscope Zeiss Axioplan I Imaging equipped with Apotome and HRm AxioCam (Zeiss, Germany) was used to document fluorescent images of tissue in sections.

I. Molecular Biological Techniques

Standard techniques such as small and large scale plasmid DNA preparations, quantification of DNA and RNA, agarose gel electrophoresis, restriction enzyme digestion, isolation of DNA fragments, ethanol precipitations, filling up reactions of 3' and 5' overhangs, dephosphorylation of 5'ends, ligation of DNA fragments, preparation and transformation of competent cells were essentially performed as described in (Sambrook J 1989).

1. Isolation, Purification and Quantification of DNA and RNA

a) Isolation of Genomic DNA

Genomic DNA from whole fish was isolated according to Hogan et al., 1986. Adult zebrafish were decapitated and viscera were removed. The tissue was frozen in liquid nitrogen and pulverized. After addition of lysis buffer (0.1 M Tris/HCl, 0.2 M NaCl, 5 mM EDTA, 0.2% w/v SDS, pH 8.5) and proteinase K (150 µg/ml) the tissue was incubated under continuous rotation at 55°C overnight. Undissolved material was pelleted. After a phenol/chloroform extraction the DNA was precipitated using 1/10 volume of 3 M NaOAc (pH5.2) and 2 volumes of 100% ethanol. The DNA was washed two times with 70% ethanol, dried and dissolved in 100-500 µl H_2O overnight at 4°C.

1.1.1. Isolation of Plasmid DNA

Various methods are available for preparing high-quality plasmid DNA. To check the identity of a clone and for sequencing, plasmid DNA was prepared in small scale. For purposes like transfection of eukaryotic cells and microinjection, plasmid DNA was prepared in large scale.

1.1.2. Small Scale Plasmid DNA Preparation (Miniprep)

In cases where the recombinant E. coli clones had to be identified, plasmid DNA was isolated in small scale. Single colonies of interest were inoculated into LB-medium (10 g Tryptone, 5 g Yeast extract, 10 g NaCl; pH 7.0) containing the appropriate antibiotics (ampicillin [100 µg/ml], tetracycline HCl [50 µg/ml], or kanamycin [50 µg/ml]) in a volume of 3 ml and grown in an orbital shaker (~300 rpm) at 37°C overnight. In general, a miniprep kit (either from Qiagen or from Sigma) was used for this purpose. The plasmid purification protocols are based on a modified alkaline lysis procedure (Birnboim and Doly, 1979), followed by binding of plasmid DNA to an anion-exchange resin under appropriate low-salt and pH conditions. RNA, proteins, and low-molecular-weight impurities are removed by a medium-salt wash. Plasmid DNA is eluted in a high-salt buffer and then concentrated and desalted by isopropanol precipitation.

1.1.3. Isolation of tissue specific RNA

The tissues used were dissected in ice-cold Phosphate Buffered Saline (PBS) and stored temporarily in RNAlater solution (Ambion). The tissue was then frozen and then mechanically dissociated using a pestle. The dissociated tissue was centrifuged for 30 sec at 5000 rpm and the supernatant was discarded. The RNA was then isolated using the RNAeasy Mini Kit of Qiagen according to the instructions of the manufactor. Genomic DNA was digested by treatment with 4.7 U of RNase-free DNase (Promega) for 30 min at 37°C.

1.1.4. Phenol/Chloroform Extraction

Reaction mixtures that had a smaller volume than 200 µl were adjusted to this volume using H_2O and phenol-chloroform extracted using an equal volume of phenol-chloroform-isoamylalcohol (PCI = 25:24:1). This mixture was vortexed and centrifuged for 1 min. The aqueous phase was carefully transferred to a new reaction tube and extracted again using 200 µl of PCI. The aqueous phase was then extracted using 200 µl chloroform. The aqueous phase was then ethanol precipitated as described in the next section.

1.1.5. Ethanol Precipitation

DNA was precipitated using 0.3 M sodium acetate and two to three volumes of ice-cold absolute ethanol. RNA was precipitated using 0.8 M lithium chloride and 2.5 volumes of ice-cold absolute ethanol. Precipitation was allowed at -20°C for 30 min or at -80°C for 10 min. After centrifugation at maximum speed for 30 min in case of DNA and 20 min in case of RNA at 4°C, the pellets were washed with 70% ethanol air-dried and re-suspended in the appropriate buffer.

1.1.6. Quantitation of DNA and RNA

The concentration of DNA and RNA in solution was estimated using agarose gel electrophoresis by comparing the intensity of the bands of interest with the 1 kb band of a 10kb ladder (DNA-Hyperladder, Bioline) of known concentration. RNA samples were denatured in 50% formamide for 3 min at 100°C before loading.

1.1.7. Agarose Gel Electrophoresis

DNA and RNA were loaded on 1% agarose gels containing 0.5 µg/ml ethidium bromide in 1 x TAE Buffer and run at 5-10 V/cm. Genomic DNA, was loaded on low concentration agarose gels (1%) gel and run slowly (1-2 V/cm) to ensure better separation and to avoid smearing the DNA. The loading dye used was purchased from Bioline. The DNA-HyperLadder I (Bioline) was used for estimation of molecular weight.

2. Enzymatic Modifications of DNA

1.1.8. Restriction enzyme digestion of DNA

Digestions for characterization of plasmid DNA were performed using about 200 ng of plasmid in 1 x restriction enzyme digestion buffer and 6 U of restriction enzyme in a total volume of 20 µl. Digestion mixtures were incubated for 1-2 h at the appropriate temperatures for each enzyme as suggested by the manufacturer

1.1.9. Ligation of DNA Fragments and PCR Products

Ligation reactions were used to combine vector and insert DNA. For this purpose purified insert DNA was ligated to dephosphorylated vector DNA using T4 DNA ligase (Roche Biochemicals) according to (Sambrook J 1989) et al., 1989 and the supplier's instructions. PCR products (1-4 µl) were ligated directly after amplification into the pBluescript II SK+ (Stratagene), pDrive (Qiagen) or pGEM-T (Promega) vector according to the manufacturer's instructions.

3. Isolation of DNA Fragments from PCR products or Agarose Gels

DNA fragments were isolated from agarose gels according to the manufacturer's instructions. In general QIAquick Gel Extraction Kit (Qiagen), QIAquick PCR Purification Kit (Qiagen) or Roche High Pure PCR Product Purification Kit (Roche Biochemicals) was used. All these kits make use of a column filled with a silica-gel membrane. DNA adsorbs to the silica-membrane in the presence of high salt while contaminants pass through the column. Impurities are washed away and the pure DNA is eluted with Tris buffer (low salt condition).

4. Labeling of RNA Using Digoxigenin, Biotin or Fluorescein by In Vitro Transcription

Various kinds of probes can be used for the detection of mRNA in *in situ* hybridization experiments. However, *in vitro* transcribed riboprobes are the best choice on tissue sections (Cox, 1984). These probes are single-stranded and may span hundreds of nucleotides, which results in specific antisense probes with high detection sensitivity. Moreover, *in vitro* transcription allows the synthesis of ideal control probes, as the sense probes have identical length and G + C content, defining similar properties of hybridization compared to the antisense probes. After synthesis of the probes they were not hydrolyzed into smaller pieces, as this treatment leads to elevated background signals. Sense and antisense RNA probes labeled with digoxigenin (DIG)-, fluorescein- or biotin-labeled UTP were generated by *in vitro* transcription according to the manufacturer's instructions (RocheBiochemicals). Before beginning the transcription reaction, the template DNA was generated either by PCR using insert specific primers that contained the T3 polymerase promoter sequence or by linearization with a restriction enzyme. For the latter, T7 or SP6 polymerase promoter sequences in the vector backbone were utilized for transctiption. The template DNA was then purified using a PCR purification kit (Qiagen). The labeling reaction was performed in a total volume of 20-40 µl. About 200-500 ng (for PCR template) or about 1000ng (for linearized plasmid template) of plasmid was used for labeling. Transcription buffer and DIG-, fluorescein- or biotin-labeling mixture were added to a final concentration of 1 x. 4 U of RNA

Polymerase (T3, T7 or SP6) and 20 U of RNAse inhibitor (Roche Biochemicals) were also added. The reaction was incubated at 38°C for 2 hours and terminated by addition of 2 µl of EDTA (200 mM, pH 8.0). The RNA transcript was ethanol precipitated and analyzed for size and integrity using agarose gel electrophoresis. Labeling efficiency was estimated using DIG quantification teststrips (Roche Biochemicals).

5. Subcloning of DNA Fragments

1.1.10. Electroporation

For electroporation the bacterial suspension of XL1 Blue strain was thawed on ice, mixed with 1-2 µl of ligation mixture and after a 1 min incubation at room temperature transformed using 1 mm cuvettes and the GenePulser from BioRad at E=18 kV/cm, C=25 µF, R=200 W. After transformation 1 ml of pre-warmed LB medium was added to the bacteria and they were incubated for 1 h at 37°C before plating on ampicillin (50 µg/ml) and tetracyclin (150 µg/ml) plates. In cases where blue-white selection was possible X-gal (800 µg/ml) and IPTG (0.5 mM) were added to the plates. The bacteria were grown on agar plates overnight at 37°C.

1.1.11. Colony PCR for Identification of Positive Clones and Determination of Insert Length

Single bacterial colonies were picked and inoculated in LB medium containing the appropriate antibiotic in 96 well multititer plates. The bacteria were grown for one to three hours in an orbital rotator at 37°C. Five µl of bacterial suspension was used as a template in a PCR reaction. All PCR reactions were carried out in a final volume of 20 µl containing 1 x PCR buffer, 1.25 mM MgCl2, 10 pmoles of each primer, 0.1 mM of each dNTP, 1 U of Taq DNA polymerase (from Roche Biochemicals). Generally M13 primers or other vector primers like T3 and T7 primers were used. Clones that were positive in the PCR were used to inoculate 3 ml of LB medium. DNA was extracted using the small-scale preparation of DNA protocol and digested to confirm the positive result of the PCR. Single clones were then subjected to sequence analysis.

1.1.12. Preparation of Glycerol Stocks

Glycerol stocks were prepared by adding 150 µl of sterile glycerol to 850 µl bacterial culture, vortexing to ensure even dispersion of the glycerol and freezing in liquid nitrogen. Afterwards, tubes were transferred to -80°C for long-term storage.

6. Synthesis of DNA

1.1.13. Polymerase Chain Reactions (PCR)

Polymerase chain reactions were performed for different purposes like subcloning of DNA fragments, identification of positive clones and determination of insert sizes in colony PCRs, preparing of a template for in vitro transcription and sequencing. Primers were designed according to standard methods using the Vector NTI software. The following conditions were used: 3 min at 96°C, followed by 35 cycles of 30 sec at 96°C, 30 sec at Tm°C, and 60 sec at 72°C, and a final extension of 10 min at 72°C. Tm was 60ºC for all the analyzed genes but *taar13c* (Tm of 50ºC).

1.1.14. Synthesis of cDNA by Reverse Transcription

The reaction mixture was prepared using 8 µl DNase-treated RNA, 15 µl DEPC treated H2O and 2.3 µl pdN6 (100 ng/µl) and incubated 10 min at 70°C and then cooled on ice for 2 min. To start the reaction 9 µl of reverse transcriptase buffer (5x Superscript II, Invitrogen Life Technologies), 2.3 µl dNTP (10 mM, Amersham) and 4.5 µl dithiothreitol (0.1 M) were added and incubated first for 10 min at 25°C and then 2 min to 42°C. The reaction was started by addition of 2.3 µl (460 U) reverse transcriptase (Superscript II, Invitrogen Life Technologies) and synthesis was allowed for 50 min at 42°C. Inactivating the reverse transcriptase for 15 min at 70°C stopped the transcription. Two µl of this product was directly used for PCR reactions.

7. Sequencing of DNA

DNA sequencing was carried at the core facility of the Institute of Genetics by Rita Lange on an ABI Prism 3730 DNA Analyzer (Applied Biosystems, USA).

For each cycle sequencing reaction, the following reagents were added into each tube: 2µl of BigDye terminator premix (ABI Prism), 3.2pmol primer, 100ng of purified plasmid DNA and autoclaved distilled water to a final volume of 10µl. Then, the mixture was mixed and briefly spun down. The sequencing profile used was as follows: 40 cycles at 95^0C for 20 sec, 50^0C for 15 sec and 60^0C for 4 min. The samples were ethanol precipitated and dried thoroughly. The dried samples were stored at -20°C in the dark until they were electrophoresed.

Sequence analysis was carried out using BLAST (Basic Local Alignment Search Tool), accessed through the Internet (http://www.ncbi.nlm.nih.gov/). Alignments of the sequences with several closely related genes were carried out subsequently.

J. Histological Studies

1. Preparation of Coverslips

To inhibit binding of antibodies and probes to coverslips these were treated with Repel Silane (Amersham Pharmacia Biotech). Coverslips were oven baked, dipped into Repel Silane, acetone and absolute ethanol for 5 sec each and air-dried in a dust-free place on the back of a microtiter plate that was used as a rack.

2. Tissue Preparation and Sectioning

Adult zebrafish were decapitated with a sharp scalpel. The head was put immediately in a petridish containing ice-cold PBS, pH 7.4. Barbels attached with lips, olfactory epithelia, olfactory bulbs, whole brains, gills, hearts and livers were dissected out.

3. Cryosectioning

For cryostat sectioning, tissues were put in TissueTek (MILES, Elkhart, Indiana, USA), oriented and frozen at -20°C. Olfactory epithelia were sectioned at 10 µm. Sections were mounted on coated Superfrost plus slides and dried for 3 h at 55°C. Sections were used immediately, since it was found that storage impaired the signals in the in situ hybridization.

K. Immunohistochemistry (IHC)

1. Antibody Staining on Fresh Frozen Cryostat Sections

Sections were fixed in 4% PFA for 10-20 min at room temperature (or alternatively overnight at 4ºC) and washed three times for 10 min each in PBS 1x (pH 7.5). Tissue in the slides was then dried by incubation in aceton for 15 min at -20ºC. The slides were washed 3 x for 5 min in PBST (PBS + 0.1% triton-100). Blocking was done in 5% normal goat serum (NGS) in PBST for at least one hour at room temperature. The tissue was then incubated with the primary antibodies in 5% NGS in PBST overnight at 4°C (or alternatively at room temperature for 2 hours). After extensive washing in PBST (3 x 10 min), the sections were incubated with the correspondent coupled Alexa-488 or -594 secondary antibodies in PBST for 2 hours at room temperature. The sections were mounted and embedded in Vectashield (Vector).

L. *In Situ* Hybridization (ISH)

In situ hybridization to cellular RNA was used to determine the cellular localization of specific messages within complex cell populations and tissues. Various methods were used for different purposes.

1. In Situ Hybridization on Sections of Olfactory Epithelia

Sections (10 □m) were fixed in 4% paraformaldehyde for 10 min at room temperature. Hybridizations were performed overnight at 60°C using standard protocols as previously described (Weth et al., 1996). Anti-DIG primary antibody coupled to alkaline phosphatase (Roche Molecular Biochemicals) and NBT-BCIP (Roche Molecular Biochemicals) were used for signal detection.

2. Fluorescent In Situ Hybridization (FISH) on Sections of Olfactory Epithelia

Hybridization with biotin-labeled probes on fresh frozen sections of olfactory epithelia were performed using the same protocol as above, with the following modifications modifications. The proteinase K step was replaced by incubation of the slides in H_2O_2 1% for 30 min at room temperature. For detection, the antibody HRP-Streptavidin (1:100 Molecular Probes, Invitrogen detection technologies) was applied first and then incubated at room temperature for 2 hours. Sections were again washed in PBST and detection of the second transcript was then performed by reaction with TSA Alexa Fluor-488 according to the manufacturer's protocol (TSA Kit, Molecular Probes, Invitrogen detection technologies). Sections were then mounted in Vectashield (containing DAPI for the counterstaining).

3. Whole Mount In Situ Hybridization of Embryos

Whole mount RNA *in situ* hybridization of zebrafish embryos or larvae was performed following the method of (Thisse, Thisse et al. 1993)as modified in Thisse and Thisse [http://zfin.org/ZFIN/Methods/ ThisseProtocol.html].

M. Double Labeling Experiments

1. In Situ Hybridization and Antibody Staining

For simultaneous detection of a transcript and a protein, first the fluorescent *in situ* hybridization was performed, since the RNA is very sensitive to degradation. The immunostaining procedure was performed on the same samples afterwards.

2. Double In Situ Hybridization

In some cases double *in situ* hybridization experiments were performed to detect transcripts of two different genes simultaneously. For this purpose, one of the RNA probes was labeled using digoxigenin or fluorescein and the other using biotin as described previously. Then *in situ* hybridization was performed by hybridizing the sample with both

probes simultaneously and using the same conditions as previously described with the following modifications. The proteinase K step was replaced by incubation of the slides in H_2O_2 1% for 30 min at room temperature. For detection, the anti-DIG- or anti-Flu-AP antibody (1:500) was applied first and incubated overnight at 4ºC. After several washing steps in PBST, the slides were treated with HNPP/ Fast Red alkaline phosphatase substrate (Roche) according to the manufacturer's instructions. The specimen were then washed 4 x 5 min with PBST, blocked (Blocking reagent from the TSA Kit, Molecular Probes, Invitrogen detection technologies) for 1 hour and incubated with HRP-Streptavidin (Molecular Probes, Invitrogen detection technologies) for 2 hours, all steps at room temperature. Sections were again washed in PBST and detection of the second transcript was then performed by reaction with TSA Alexa Fluor-488 according to the manufacturer's protocol (TSA Kit, Molecular Probes, Invitrogen detection technologies). Sections were then mounted in Vectashield (containing DAPI for the counterstaining).

All the sections that underwent double-label *in situ* hybridization were observed and photographs were taken under a fluorescent microscope (Axioplan I Imaging with Apotome, Zeiss, Germany) with sequential imaging of the two channels. All photographs in figures are single plane images (1-2 um optical section thickness).

N. Data mining

1. V1r

All annotated V1R sequences were extracted either from the National Center for Biotechnology Information (NCBI) database resources or from the articles in which they were first published (for cow, dog, frog, fish and opossum genes) and used as query sequences in subsequent analysis. Two combined different strategies were used to search the databases for new V1R-like candidate genes in five fish species and in frog. First, the algorithm tBLASTN was applied to compare amino-acid query sequences to the DNA databases (http://www.ensembl.org/index.html) with a nonstringent expectation cut-off value of 10–10; second and last, the automatically ortholog predicted genes in the fish species were retrieved from each of the query sequences. (Hubbard, Aken et al. 2007). For shark and lamprey, also EST databanks were searched, in addition, for elephant shark (*Callorhinchus milii*) WGS sequences with 1.4 fold genomic coverage were analysed.

To be considered as validated *ora*/V1R-like genes, a triage of the candidates was performed using different inclusion criteria. The confirmed *ora*/V1R-like genes were then included in subsequent analysis as new query sequences, until no new candidates were found.

The inclusion criteria used were: a) position within the *Ora*/V1R-like clade in the phylogenetic analysis; b) application of the BLASTP algorithm in the NCBI non-redundant database should result in annotated V1Rs or some other *ora*/V1R-like candidates as first hits (expectation cut-off value of 10–10); c) presence of typical V1R family motifs; d) CDS length between 850 and 1250 amino acids; e) presence of seven trans-membrane domains (based on the consensus of the prediction results obtained by using the TMHMM: http://www.cbs.dtu.dk/services/TMHMM/ and TMpred: http://www.ch.embnet.org/software/TMPRED_form.html servers).

Duplicates genes were removed and the resulting genes were subjected to the analyses described below. The *Xenopus tropicalis* automatically annotated orthologs of the fish *ora* genes and mammalian V1Rs were also retrieved and included in the subsequent phylogenetic analysis. For Genbank accession numbers see Figure 32. Sequenced zebrafish nucleotide *ora* fragments EMBL accession numbers: *ora1* - AM778163, *ora2* - AM778164, *ora3* - AM778165, *ora4* - AM778166, *ora5* - AM778167, *ora6* - AM778168.

2. Taar

All annotated TAAR sequences were compiled and used as query in tblastN searches in the NCBI and Ensembl databanks. Additionally blastP searches were performed in the NCBI databanks and automated ortholog prediction was used in the Ensembl databank (Hubbard, Aken et al. 2007). For shark, lamprey and zebrafish, also EST databanks were searched, in addition, for elephant shark WGS sequences with 1.4 fold genomic coverage were analysed. Search was recursive until no new candidates were found. Validation of candidates as proper *taar* genes required: a) position within the TAAR clade in the phylogenetic analysis; b) application of the BLASTP algorithm in the NCBI nonredundant database should result in confirmed TAARs as first hits; c) presence of typical TAAR family motifs; d) CDS length between 800 and 1300 amino acids; e) presence of seven trans-membrane domains (regions assignment according to conserved position as described in (Lindemann and Hoener 2005). For accession numbers of the genes cloned and used in the experiments see Fig. 32. For the accession numbers of the remaining *taar* genes see (Hussain, Saraiva et al. 2008).

O. Phylogenetic analysis

MAFFT, online version 5.8 (http://align.bmr.kyushu-u.ac.jp/mafft/online/server/), was employed for multiple protein alignments using the E-INS-i strategy with the default parameters. To estimate the phylogenetic relationships of the sequences we performed distance-based, maximum parsimony, and maximum likelihood analyses using the Neighbor Joining (NJ), Protpars (MP) and Proml (ML) programs as implemented in ClustalX [53] and

PHYLIP [22], packages respectively. For the NJ method I performed bootstrapping with 1000 repetitions using ClustalX [53] and for the MP and ML methods I performed bootstrapping with 1000 and 100 repetitions (for the *ora* and *taar* gene families, respectively) using the program SEQBOOT from the PHYLIP package [22]. The three methods gave similar clustering. Consensus trees were obtained using the CONSENSE program of the PHYLIP package [22].

P. dN/dS analysis

The global *d*N/*d*S ratios for the full-length ORF of the 28 fish *ora* receptor coding sequences were determined using both the DnaSP 4.10 software package (Rozas et al. 2003), which implements previously published methods (Nei and Gojobori 1986) and the Single Likelihood Ancestor Counting (SLAC) package (http://www.datamonkey.org), which implements the Suzuki-Gojobori method (Suzuki and Gojobori 1999). Since the values obtained by both methods did not vary, for the *taar* genes only the mentioned SLAC package was used. The nucleotide alignments were manually edited to match the amino acid alignment used in the phylogenetic trees and sequence logo.

To make inferences about selective pressure (positive and negative selection) on individual codons (sites) within the coding sequence of the teleost fish *ora* and *taar* genes, the Single Likelihood Ancestor Counting (SLAC) package (http://www.datamonkey.org), which implements the Suzuki-Gojobori method (Suzuki and Gojobori 1999), was used.

The algorithm is briefly outlined. First, a best-fitting nucleotide substitution model was automatically selected by fitting several such substitution models to both the data and a neighbor-joining tree generated from the alignment described above. Taking the obtained substitution rates and branch lengths as constant, a codon model was employed to fit to the data and a global dN/dS ratio was calculated. Then a codon by codon reconstruction of the ancestral sequences was performed using maximum likelihood. Afterwards the expected normalized (ES) and observed numbers (EN) of synonymous (NS) and non-synonymous (NN) substitutions were calculated for each non-constant site. dN = NN/EN and dS = NS/ES were then computed, and if dN < dS (negative selection) or dN > dS (positive selection), a p-value derived from a two-tailed extended binomial distribution was used to assess significance. Tests on simulated data (S.L.K. Pond and S.D.W. Frost, methods available at http://www.datamonkey.org) show that p values equal or smaller than 0.1 identify nearly all true positives with a false positive rate generally below the nominal p value; for actual data, the number of true positives at a given false positive rate is lower.

In the present study, two thresholds for significance (0.1 and 0.2) were taken into account in order to identify residues potentially involved in odorant-binding activities.

Q. In silico prediction of *ora* coupling specificity to G-proteins

Predicitons were performed using the PredCouple2 software package (http://athina.biol.uoa.gr/bioinformatics/PRED-COUPLE2/) ((Sgourakis, Bagos et al. 2005; Sgourakis, Bagos et al. 2005)). The 28 ora amino acids sequences were used as query. The method used is based on a refined library of highly-discriminative Hidden Markov Models. Hits from individual profiles are combined by a feed-forward Artificial Neural Network to produce the final output. Seven (7) transmembrane receptor signatures from the Pfam database version 17.00 are also applied in order to verify a true GPCR sequence. When a query sequence is not recognized as a 7 transmembrane receptor a message is shown. In the present study, two thresholds for significance (90 and 95%) were taken into account in order to consider a prediction to be valid. In case two different families were predicted for the same amino acid sequence, the one with the highest value was taken. Sequences with prediction values with a threshold for significance lower than 90% were listed as “non-pred”.

X. REFERENCES

Abbasi, A. A., Z. Paparidis, et al. (2007). "Human GLI3 Intragenic Conserved Non-Coding Sequences Are Tissue-Specific Enhancers." PLoS ONE **2**(2): e366.

Abogadie, F. C., R. C. Bruch, et al. (1995). "G-protein subunits expressed in catfish olfactory receptor neurons." Chem Senses **20**(2): 199-206.

Abogadie, F. C., R. C. Bruch, et al. (1995). "Molecular cloning of a phosphoinositide-specific phospholipase C from catfish olfactory rosettes." Brain Res Mol Brain Res **31**(1-2): 10-6.

Ache, B. W. and J. M. Young (2005). "Olfaction: diverse species, conserved principles." Neuron **48**(3): 417-30.

Alioto, T. S. and J. Ngai (2005). "The odorant receptor repertoire of teleost fish." BMC Genomics **6**(173): 173.

Alioto, T. S. and J. Ngai (2006). "The repertoire of olfactory C family G protein-coupled receptors in zebrafish: candidate chemosensory receptors for amino acids." BMC Genomics **7**(309): 309.

Anholt, R. R. and A. M. Rivers (1990). "Olfactory transduction: cross-talk between second-messenger systems." Biochemistry **29**(17): 4049-54.

Axel, R. (1995). "The molecular logic of smell." Sci Am **273**(4): 154-9.

Bastianelli, E. and R. Pochet (1995). "Calbindin-D28k, calretinin, and S-100 immunoreactivities in rat pineal gland during postnatal development." J Pineal Res **18**(3): 127-34.

Bastianelli, E. and R. Pochet (1995). "Calmodulin, calbindin-D28k, calretinin and neurocalcin in rat olfactory bulb during postnatal development." Brain Res Dev Brain Res **87**(2): 224-7.

Bastianelli, E., A. S. Polans, et al. (1995). "Differential distribution of six calcium-binding proteins in the rat olfactory epithelium during postnatal development and adulthood." J Comp Neurol **354**(3): 395-409.

Bastianelli, E., K. Takamatsu, et al. (1995). "Hippocalcin in rat retina. Comparison with calbindin-D28k, calretinin and neurocalcin." Exp Eye Res **60**(3): 257-66.

Beites, C. L., S. Kawauchi, et al. (2005). "Identification and molecular regulation of neural stem cells in the olfactory epithelium." Exp Cell Res **306**(2): 309-16.

Belanger, R. M., C. M. Smith, et al. (2003). "Morphology and histochemistry of the peripheral olfactory organ in the round goby, Neogobius melanostomus (Teleostei: Gobiidae)." J Morphol **257**(1): 62-71.

Berghard, A. and L. B. Buck (1996). "Sensory transduction in vomeronasal neurons: evidence for G alpha o, G alpha i2, and adenylyl cyclase II as major components of a pheromone signaling cascade." J Neurosci **16**(3): 909-18.

Berghard, A., L. B. Buck, et al. (1996). "Evidence for distinct signaling mechanisms in two mammalian olfactory sense organs." Proc Natl Acad Sci U S A **93**(6): 2365-9.

Biffo, S., T. Goren, et al. (1991). "Expression of calmodulin mRNA in rat olfactory neuroepithelium." Brain Res Mol Brain Res **10**(1): 13-21.

Borowsky, B., N. Adham, et al. (2001). "Trace amines: identification of a family of mammalian G protein-coupled receptors." Proc Natl Acad Sci U S A **98**(16): 8966-71.

Boschat, C., C. Pelofi, et al. (2002). "Pheromone Detection Mediated by a V1r Vomeronasal Receptor." Nat Neurosci. **5**(12): 1261-2.

Bryson-Richardson, R. J., D. W. Logan, et al. (2004). "Large-scale analysis of gene structure in rhodopsin-like GPCRs: evidence for widespread loss of an ancient intron." Gene **338**(1): 15-23.

Buck, L. and R. Axel (1991). "A novel multigene family may encode odorant receptors: a molecular basis for odor recognition." Cell **65**(1): 175-87.

Buck, L. B. (1996). "Information coding in the vertebrate olfactory system." Annu Rev Neurosci **19**: 517-44.

Buck, L. B. (2000). "The molecular architecture of odor and pheromone sensing in mammals." Cell **100**(6): 611-8.
Cao, Z., R. L. Tanguay, et al. (2003). "Identification of a putative calcium-binding protein as a dioxin-responsive gene in zebrafish and rainbow trout." Aquat Toxicol **63**(3): 271-82.
Carmel, L., I. B. Rogozin, et al. (2007). "Evolutionarily conserved genes preferentially accumulate introns." Genome Res **10**: 10.
Carmel, L., Y. I. Wolf, et al. (2007). "Three distinct modes of intron dynamics in the evolution of eukaryotes." Genome Res **10**: 10.
Celik, A., S. H. Fuss, et al. (2002). "Selective targeting of zebrafish olfactory receptor neurons by the endogenous OMP promoter." Eur J Neurosci **15**(5): 798-806.
Date-Ito, A., H. Ohara, et al. (2008). "Xenopus V1R Vomeronasal Receptor Family Is Expressed in the Main Olfactory System." Chem Senses.
Del Punta, K., A. Rothman, et al. (2000). "Sequence diversity and genomic organization of vomeronasal receptor genes in the mouse." Genome Res **10**(12): 1958-67.
Dellacorte, C., D. Restrepo, et al. (1996). "G alpha 9/G alpha 11: immunolocalization in the olfactory epithelium of the rat (Rattus rattus) and the channel catfish (Ictalurus punctatus)." Neuroscience **74**(1): 261-73.
Doving, K. B. and R. Selset (1980). "Behavior patterns in cod released by electrical stimulation of olfactory tract bundlets." Science **207**(4430): 559-60.
Dryer, L. (2000). "Evolution of odorant receptors." Bioessays **22**(9): 803-10.
Duda, T. and R. K. Sharma (2004). "S100B-modulated Ca2+-dependent ROS-GC1 transduction machinery in the gustatory epithelium: a new mechanism in gustatory transduction." FEBS Lett **577**(3): 393-8.
Dulac, C. (2000). "Sensory coding of pheromone signals in mammals." Curr Opin Neurobiol **10**(4): 511-8.
Dulac, C. and R. Axel (1995). "A novel family of genes encoding putative pheromone receptors in mammals." Cell **83**(2): 195-206.
Elsaesser, R. and J. Paysan (2005). "Morituri te salutant? Olfactory signal transduction and the role of phosphoinositides." J Neurocytol **34**(1-2): 97-116.
Elsaesser, R. and J. Paysan (2007). "The sense of smell, its signalling pathways, and the dichotomy of cilia and microvilli in olfactory sensory cells." BMC Neurosci **8 Suppl 3**: S1.
Eriksson, N. E. (1984). "Clustering of foodstuffs in food hypersensitivity. An inquiry study in pollen allergic patients." Allergol Immunopathol (Madr) **12**(1): 28-32.
Feinstein, P., T. Bozza, et al. (2004). "Axon guidance of mouse olfactory sensory neurons by odorant receptors and the beta2 adrenergic receptor." Cell **117**(6): 833-46.
Feinstein, P. and P. Mombaerts (2004). "A contextual model for axonal sorting into glomeruli in the mouse olfactory system." Cell **117**(6): 817-31.
Ferrando, S., M. Bottaroa, et al. (2007). "First detection of olfactory marker protein (OMP) immunoreactivity in the olfactory epithelium of a cartilaginous fish." Neuroscience Letters **413**(2): 173-176.
Ferrando, S., M. Bottaroa, et al. (2007). "Appearance of Crypt Neurons in the Olfactory Epithelium of the Skate (Raja calvata) Durung Development." The Antomical Record **290**: 1268-1272.
Firestein, S. (1992). "Electrical signals in olfactory transduction." Curr Opin Neurobiol **2**(4): 444-8.
Firestein, S. (1996). "Olfaction: scents and sensibility." Curr Biol **6**(6): 666-7.
Firestein, S., H. Breer, et al. (1996). "Olfaction: what's new in the nose." J Neurobiol **30**(1): 1-2.
Fredriksson, R., M. C. Lagerstrom, et al. (2003). "The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints." Mol Pharmacol **63**(6): 1256-72.
Fridmanis, D., R. Fredriksson, et al. (2007). "Formation of new genes explains lower intron density in Rhodopsin G protein-coupled receptors." Mol Phylogenet Evol **43**(3): 864-80.

Friedrich, R. W. and S. I. Korsching (1997). "Combinatorial and chemotopic odorant coding in the zebrafish olfactory bulb visualized by optical imaging." Neuron **18**(5): 737-52.

Friedrich, R. W. and S. I. Korsching (1998). "Chemotopic, Combinatorial, and Noncombinatorial Odorant Representations in the Olfactory Bulb Revealed Using a Voltage-Sensitive Axon Tracer." J Neurosci. **18**(23): 9977-88.

Germana, A., G. Montalbano, et al. (2004). "S100 protein-like immunoreactivity in the crypt olfactory neurons of the adult zebrafish." Neurosci Lett **371**(2-3): 196-8.

Gibbs, R. A., G. M. Weinstock, et al. (2004). "Genome sequence of the Brown Norway rat yields insights into mammalian evolution." Nature **428**(6982): 493-521.

Gilad, Y., D. Segre, et al. (2000). "Dichotomy of single-nucleotide polymorphism haplotypes in olfactory receptor genes and pseudogenes." Nat Genet **26**(2): 221-4.

Gloriam, D. E., T. K. Bjarnadottir, et al. (2005). "The repertoire of trace amine G-protein-coupled receptors: large expansion in zebrafish." Mol Phylogenet Evol **35**(2): 470-82.

Gojobori, T. (1983). "Codon substitution in evolution and the "saturation" of synonymous changes." Genetics **105**(4): 1011-27.

Gold, G. H. (1999). "Controversial issues in vertebrate olfactory transduction." Annu Rev Physiol **61**: 857-71.

Grus, W. E., P. Shi, et al. (2005). "Dramatic variation of the vomeronasal pheromone receptor gene repertoire among five orders of placental and marsupial mammals." Proc Natl Acad Sci U S A **102**(16): 5767-72.

Hamdani, E. H., G. Alexander, et al. (2001). "Projection of sensory neurons with microvilli to the lateral olfactory tract indicates their participation in feeding behaviour in crucian carp." Chem Senses **26**(9): 1139-44.

Hamdani, E. H., A. Kasumyan, et al. (2001). "Is feeding behaviour in crucian carp mediated by the lateral olfactory tract?" Chem Senses **26**(9): 1133-8.

Hamdani, E. H., S. Lastein, et al. (2007). "Seasonal Variations in Olfactory Sensory Neurons Fish Sensitivity to Sex Pheromones Explained?" Chem Senses **24**: 24.

Hamdani, E. H., O. B. Stabell, et al. (2000). "Alarm reaction in the crucian carp is mediated by the medial bundle of the medial olfactory tract." Chem Senses **25**(1): 103-9.

Hamdani el, H. and K. B. Doving (2002). "The alarm reaction in crucian carp is mediated by olfactory neurons with long dendrites." Chem Senses **27**(4): 395-8.

Hamdani el, H. and K. B. Doving (2006). "Specific projection of the sensory crypt cells in the olfactory system in crucian carp, Carassius carassius." Chem Senses **31**(1): 63-7.

Hansen, A. (2006). "Unconventional sensory cells in the nasal epithelia of rodents and humans " Chem Senses **31**: E4.

Hansen, A., K. T. Anderson, et al. (2004). "Differential distribution of olfactory receptor neurons in goldfish: structural and molecular correlates." J Comp Neurol **477**(4): 347-59.

Hansen, A. and T. E. Finger (2000). "Phyletic distribution of crypt-type olfactory receptor neurons in fishes." Brain Behav Evol **55**(2): 100-10.

Hansen, A., S. H. Rolen, et al. (2003). "Correlation between olfactory receptor cell type and function in the channel catfish." J Neurosci **23**(28): 9328-39.

Hansen, A. and M. Schmidt (2004). "Influence of season and environment on adult neurogenesis in the central olfactory pathway of the shore crab, Carcinus maenas." Brain Res **1025**(1-2): 85-97.

Hansen, A. and E. Zeiske (1998). "The peripheral olfactory organ of the zebrafish, Danio rerio: an ultrastructural study." Chem Senses **23**(1): 39-48.

Hansen, A. and B. S. Zielinski (2005). "Diversity in the olfactory epithelium of bony fishes: development, lamellar arrangement, sensory neuron cell types and transduction components." J Neurocytol **34**(3-5): 183-208.

Harden, M. V., L. A. Newton, et al. (2006). "Olfactory imprinting is correlated with changes in gene expression in the olfactory epithelia of zebrafish." Journal of Neurobiology **66**(13): 1452-1466.

Hashiguchi, Y. and M. Nishida (2005). "Evolution of Vomeronasal-Type Odorant Receptor Genes in the Zebrafish Genome." Gene **362**: 19-28. Epub 2005 Oct 14.

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

Hashiguchi, Y. and M. Nishida (2007). "Evolution of trace amine associated receptor (TAAR) gene family in vertebrates: lineage-specific expansions and degradations of a second class of vertebrate chemosensory receptors expressed in the olfactory epithelium." Mol Biol Evol **24**(9): 2099-107.

Heizmann, C. W. and J. A. Cox (1998). "New perspectives on S100 proteins: a multi-functional Ca(2+)-, Zn(2+)- and Cu(2+)-binding protein family." BioMetals **11**(4): 383-97.

Hermans, E. and R. A. Challiss (2001). "Structural, signalling and regulatory properties of the group I metabotropic glutamate receptors: prototypic family C G-protein-coupled receptors." Biochem J **359**(Pt 3): 465-84.

Herrada, G. and C. Dulac (1997). "A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution." Cell **90**(4): 763-73.

Hildebrand, J. G. and G. M. Shepherd (1997). "Mechanisms of olfactory discrimination: converging evidence for common principles across phyla." Annu Rev Neurosci **20**: 595-631.

Hoegg, S., H. Brinkmann, et al. (2004). "Phylogenetic timing of the fish-specific genome duplication with the diversification of teleost fish." J Mol Evol **59**(2): 190-203.

Huang, E. S. (2003). "Construction of a sequence motif characteristic of aminergic G protein-coupled receptors." Protein Sci **12**(7): 1360-7.

Hubbard, T. J., B. L. Aken, et al. (2007). "Ensembl 2007." Nucleic Acids Res **35**(Database issue): D610-7.

Hughes, A. L. and M. K. Hughes (1993). "Adaptive evolution in the rat olfactory receptor gene family." J Mol Evol **36**(3): 249-54.

Hussain, A., L. R. Saraiva, et al. (2008). "Strong positive selection and high intron dynamics in teleost taar genes." PNAS, under review.

Ishimaru, Y., S. Okada, et al. (2005). "Two families of candidate taste receptors in fishes." Mech Dev **122**(12): 1310-21.

Kraemer, A. M., L. R. Saraiva, et al. (2008). "Structural and functional diversification in the teleost S100 family of calcium-binding proteins." BMC Evol Biol **8**(48): 48.

Krieger, J. and H. Breer (1999). "Olfactory reception in invertebrates." Science **286**(5440): 720-3.

Kunishima, N., Y. Shimada, et al. (2000). "Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor." Nature **407**(6807): 971-7.

Lane, R. P., J. Young, et al. (2004). "Species specificity in rodent pheromone receptor repertoires." Genome Res **14**(4): 603-8.

Lastein, S., H. Hamdani el, et al. (2006). "Gender distinction in neural discrimination of olfactory bulb of crucian carp, Carassius." Chem Senses **31**(1): 69-77.

Lewcock, J. W. and R. R. Reed (2004). "A feedback mechanism regulates monoallelic odorant receptor expression." Proc Natl Acad Sci U S A **101**(4): 1069-74.

Liberles, S. D. and L. B. Buck (2006). "A second class of chemosensory receptors in the olfactory epithelium." Nature **442**(7103): 645-50.

Lindemann, L., M. Ebeling, et al. (2005). "Trace amine-associated receptors form structurally and functionally distinct subfamilies of novel G protein-coupled receptors." Genomics **85**(3): 372-85.

Lindemann, L. and M. C. Hoener (2005). "A renaissance in trace amines inspired by a novel GPCR family." Trends Pharmacol Sci **26**(5): 274-81.

Lo, Y. H., T. M. Bradley, et al. (1993). "Stimulation of Ca(2+)-regulated olfactory phospholipase C by amino acids." Biochemistry **32**(46): 12358-62.

Ma, L. and W. C. Michel (1998). "Drugs affecting phospholipase C-mediated signal transduction block the olfactory cyclic nucleotide-gated current of adult zebrafish." J Neurophysiol **79**(3): 1183-92.

Mammen, A., P. J. Simpson, et al. (2004). "Hippocalcin in the olfactory epithelium: a mediator of second messenger signaling." Biochem Biophys Res Commun **322**(4): 1131-9.

Mank, J. E., E. Axelsson, et al. (2007). "Fast-X on the Z: rapid evolution of sex-linked genes in birds." Genome Res **17**(5): 618-24.

Marenholz, I., C. W. Heizmann, et al. (2004). "S100 proteins in mouse and man: from evolution to function and pathology (including an update of the nomenclature)." Biochem Biophys Res Commun **322**(4): 1111-22.

Martini, S., L. Silvotti, et al. (2001). "Co-expression of putative pheromone receptors in the sensory neurons of the vomeronasal organ." J Neurosci **21**(3): 843-8.

Matsunami, H. and L. B. Buck (1997). "A multigene family encoding a diverse array of putative pheromone receptors in mammals." Cell **90**(4): 775-84.

Matthews, H. R. and J. Reisert (2003). "Calcium, the two-faced messenger of olfactory transduction and adaptation." Curr Opin Neurobiol **13**(4): 469-75.

Menini, A. (1999). "Calcium signalling and regulation in olfactory neurons." Curr Opin Neurobiol **9**(4): 419-26.

Meyer, A. and M. Schartl (1999). "Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions." Curr Opin Cell Biol **11**(6): 699-704.

Miwa, N., M. Kobayashi, et al. (1998). "Purification and molecular cloning of a novel calcium-binding protein, p26olf, in the frog olfactory epithelium." Biochem Biophys Res Commun **251**(3): 860-7.

Miyamoto, T., D. Restrepo, et al. (1992). "IP3- and cAMP-induced responses in isolated olfactory receptor neurons from the channel catfish." J Membr Biol **127**(3): 173-83.

Miyamoto, T., D. Restrepo, et al. (1992). "Voltage-dependent and odorant-regulated currents in isolated olfactory receptor neurons of the channel catfish." J Gen Physiol **99**(4): 505-29.

Mombaerts, P. (2004). "Genes and ligands for odorant, vomeronasal and taste receptors." Nat Rev Neurosci **5**(4): 263-78.

Moncada, S. and J. P. Bolanos (2006). "Nitric oxide, cell bioenergetics and neurodegeneration." J Neurochem **97**(6): 1676-89.

Mori, I. (1999). "Genetics of chemotaxis and thermotaxis in the nematode Caenorhabditis elegans." Annu Rev Genet **33**: 399-422.

Morita, Y. and T. E. Finger (1998). "Differential projections of ciliated and microvillous olfactory receptor cells in the catfish, Ictalurus punctatus." J Comp Neurol **398**(4): 539-50.

Murphy, W. J., T. H. Pringle, et al. (2007). "Using genomic data to unravel the root of the placental mammal phylogeny." Genome Res **17**(4): 413-21.

Nakamura, T. (2000). "Cellular and molecular constituents of olfactory sensation in vertebrates." Comp Biochem Physiol A Mol Integr Physiol **126**(1): 17-32.

Nakatani, Y., H. Takeda, et al. (2007). "Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates." Genome Res **17**(9): 1254-65.

Nei, M. and T. Gojobori (1986). "Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions." Mol Biol Evol **3**(5): 418-26.

Ngai, J., A. Chess, et al. (1993). "Coding of olfactory information: topography of odorant receptor expression in the catfish olfactory epithelium." Cell **72**(5): 667-80.

Ngai, J., M. M. Dowling, et al. (1993). "The family of genes encoding odorant receptors in the channel catfish." Cell **72**(5): 657-66.

Niimura, Y. and M. Nei (2003). "Evolution of olfactory receptor genes in the human genome." Proc Natl Acad Sci U S A **100**(21): 12235-40.

Niimura, Y. and M. Nei (2005). "Evolutionary Dynamics of Olfactory Receptor Genes in Fishes and Tetrapods." Proc Natl Acad Sci U S A. **102**(17): 6039-44. Epub 2005 Apr 11.

Niimura, Y. and M. Nei (2006). "Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates." J Hum Genet **51**(6): 505-17.
Niimura, Y. and M. Nei (2007). "Extensive gains and losses of olfactory receptor genes in Mammalian evolution." PLoS ONE **2**(1): e708.
Nikonov, A. A. and J. Caprio (2001). "Electrophysiological evidence for a chemotopy of biologically relevant odors in the olfactory bulb of the channel catfish." J Neurophysiol **86**(4): 1869-76.
Pantages, E. and C. Dulac (2000). "A novel family of candidate pheromone receptors in mammals." Neuron **28**(3): 835-45.
Pfister, P., J. Randall, et al. (2007). "Divergent evolution among teleost V1r receptor genes." PLoS ONE **2**(4): e379.
Pfister, P. and I. Rodriguez (2005). "Olfactory Expression of a Single and Highly Variable V1r Pheromone Receptor-Like Gene in Fish Species." Proc Natl Acad Sci U S A. **102**(15): 5489-94. Epub 2005 Apr 4.
Rawson, N. E. and A. S. LaMantia (2007). "A speculative essay on retinoic acid regulation of neural stem cells in the developing and aging olfactory system." Exp Gerontol **42**(1-2): 46-53.
Rebecchi, M. J. and S. N. Pentyala (2000). "Structure, function, and control of phosphoinositide-specific phospholipase C." Physiol Rev **80**(4): 1291-335.
Reed, R. R. (1992). "Signaling pathways in odorant detection." Neuron **8**(2): 205-9.
Reisert, J., P. J. Bauer, et al. (2003). "The Ca-activated Cl channel and its control in rat olfactory receptor neurons." J Gen Physiol **122**(3): 349-63.
Restrepo, D., I. Boekhoff, et al. (1993). "Rapid kinetic measurements of second messenger formation in olfactory cilia from channel catfish." Am J Physiol **264**(4 Pt 1): C906-11.
Restrepo, D., Y. Okada, et al. (1993). "Odorant-regulated Ca2+ gradients in rat olfactory neurons." J Gen Physiol **102**(5): 907-24.
Restrepo, D., Y. Okada, et al. (1993). "Human olfactory neurons respond to odor stimuli with an increase in cytoplasmic Ca2+." Biophys J **64**(6): 1961-6.
Restrepo, D., J. H. Teeter, et al. (1996). "Second messenger signaling in olfactory transduction." J Neurobiol **30**(1): 37-48.
Robertson, H. M. (1998). "Two large families of chemoreceptor genes in the nematodes Caenorhabditis elegans and Caenorhabditis briggsae reveal extensive gene duplication, diversification, movement, and intron loss." Genome Res **8**(5): 449-63.
Rodriguez, I., K. Del Punta, et al. (2002). "Multiple new and isolated families within the mouse superfamily of V1r vomeronasal receptors." Nat Neurosci **5**(2): 134-40.
Rodriguez, I., P. Feinstein, et al. (1999). "Variable patterns of axonal projections of sensory neurons in the mouse vomeronasal system." Cell **97**(2): 199-208.
Rolen, S. H., P. W. Sorensen, et al. (2003). "Polyamines as olfactory stimuli in the goldfish Carassius auratus." J Exp Biol **206**(Pt 10): 1683-96.
Roppolo, D., S. Vollery, et al. (2007). "Gene cluster lock after pheromone receptor gene choice." Embo J **26**(14): 3423-30.
Runnenburger, K., H. Breer, et al. (2002). "Selective G protein beta gamma-subunit compositions mediate phospholipase C activation in the vomeronasal organ." Eur J Cell Biol **81**(10): 539-47.
Ryba, N. J. and R. Tirindelli (1997). "A new multigene family of putative pheromone receptors." Neuron **19**(2): 371-9.
Saito, H., M. L. Mimmack, et al. (1998). "Isolation of mouse vomeronasal receptor genes and their co-localization with specific G-protein messenger RNAs." Brain Res Mol Brain Res **60**(2): 215-27.
Sambrook J, M., T., and Fritsch, E. F. I. (1989). Molecular Cloning: A Laboratory Manual. NY, Cold Spring Harbor Laboratory Press.
Santamaria-Kisiel, L., A. C. Rintala-Dempsey, et al. (2006). "Calcium-dependent and -independent interactions of the S100 protein family." Biochem J **396**(2): 201-14.
Saraiva, L. R. and S. I. Korsching (2007). "A novel olfactory receptor gene family in teleost fish." Genome Res **17**(10): 1448-57.

Sato, Y., N. Miyasaka, et al. (2005). "Mutually exclusive glomerular innervation by two distinct types of olfactory sensory neurons revealed in transgenic zebrafish." J Neurosci **25**(20): 4889-97.

Sato, Y., N. Miyasaka, et al. (2007). "Hierarchical regulation of odorant receptor gene choice and subsequent axonal projection of olfactory sensory neurons in zebrafish." J Neurosci **27**(7): 1606-15.

Schild, D. and D. Restrepo (1998). "Transduction mechanisms in vertebrate olfactory receptor cells." Physiol Rev **78**(2): 429-66.

Schioth, H. B. and R. Fredriksson (2005). "The GRAFS classification system of G-protein coupled receptors in comparative perspective." Gen Comp Endocrinol **142**(1-2): 94-101.

Schmachtenberg, O. (2006). "Histological and electrophysiological properties of crypt cells from the olfactory epithelium of the marine teleost Trachurus symmetricus." J Comp Neurol **495**(1): 113-21.

Schwob, J. E. (2002). "Neural regeneration and the peripheral olfactory system." Anat Rec **269**(1): 33-49.

Serizawa, S., K. Miyamichi, et al. (2003). "Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse." Science **302**(5653): 2088-94.

Sgourakis, N. G., P. G. Bagos, et al. (2005). "Prediction of the coupling specificity of GPCRs to four families of G-proteins using hidden Markov models and artificial neural networks." Bioinformatics **21**(22): 4101-6.

Sgourakis, N. G., P. G. Bagos, et al. (2005). "A method for the prediction of GPCRs coupling specificity to G-proteins using refined profile Hidden Markov Models." BMC Bioinformatics **6**: 104.

Shi, P. and J. Zhang (2007). "Comparative genomic analysis identifies an evolutionary shift of vomeronasal receptor gene repertoires in the vertebrate transition from water to land." Genome Res **17**(2): 166-74.

Shykind, B. M., S. C. Rohani, et al. (2004). "Gene switching and the stability of odorant receptor gene choice." Cell **117**(6): 801-15.

Sitaramayya, A. (2002). "Calcium-dependent activation of guanylate cyclase by S100b." Adv Exp Med Biol **514**: 389-98.

Speca, D. J., D. M. Lin, et al. (1999). "Functional identification of a goldfish odorant receptor." Neuron **23**(3): 487-98.

Spehr, M., G. Gisselmann, et al. (2003). "Identification of a testicular odorant receptor mediating human sperm chemotaxis." Science **299**(5615): 2054-8.

Stacey, N., A. Chojnacki, et al. (2003). "Hormonally derived sex pheromones in fish: exogenous cues and signals from gonad to brain." Can J Physiol Pharmacol **81**(4): 329-41.

Sumiyama, K., S. Q. Irvine, et al. (2002). "Genomic structure and functional control of the Dlx3-7 bigene cluster." Proc Natl Acad Sci U S A **99**(2): 780-5.

Thisse, C., B. Thisse, et al. (1993). "Structure of the zebrafish snail1 gene and its expression in wild-type, spadetail and no tail mutant embryos." Development **119**(4): 1203-15.

Van de Peer, Y. (2004). "Tetraodon genome confirms Takifugu findings: most fish are ancient polyploids." Genome Biol **5**(12): 250.

Vanderhaeghen, P., S. Schurmans, et al. (1993). "Olfactory receptors are displayed on dog mature sperm cells." J Cell Biol **123**(6 Pt 1): 1441-52.

Venkatesh, B., M. V. Erdmann, et al. (2001). "Molecular synapomorphies resolve evolutionary relationships of extant jawed vertebrates." Proc Natl Acad Sci U S A **98**(20): 11382-7.

von Rekowski, C. and H. P. Zippel (1993). "In goldfish the qualitative discriminative ability for odors rapidly returns after bilateral nerve axotomy and lateral olfactory tract transection." Brain Res **618**(2): 338-40.

Waterston, R. H., E. S. Lander, et al. (2002). "On the sequencing of the human genome." Proc Natl Acad Sci U S A **99**(6): 3712-6.

Weltzien, F. A., E. Hoglund, et al. (2003). "Does the lateral bundle of the medial olfactory tract mediate reproductive behavior in male crucian carp?" Chem Senses **28**(4): 293-300.

Weth, F. (2001). Molekulare Aspekte der funktionellen Architektur des Geruchssystems beim Zebrabärbling, Danio rerio. Fakultät für Biologie. Tübingen, Eberhard-Karls-Universität. **Dr. rer. nat.:** 240.

Weth, F., W. Nadler, et al. (1996). "Nested expression domains for odorant receptors in zebrafish olfactory epithelium." Proc Natl Acad Sci U S A **93**(23): 13321-6.

Wolfe, K. H. and P. M. Sharp (1993). "Mammalian gene evolution: nucleotide sequence divergence between mouse and rat." J Mol Evol **37**(4): 441-56.

Yamamoto, Y., M. Mori, et al. (2004). "Differential expression of histochemical characteristics in the developing olfactory receptor cells in a flatfish, barfin flounder (Verasper moseri)." J Vet Med Sci **66**(12): 1609-11.

Yang, H., P. Shi, et al. (2005). "Composition and Evolution of the V2r Vomeronasal Receptor Gene Repertoire in Mice and Rats." Genomics. **86**(3): 306-15.

Young, J. M., C. Friedman, et al. (2002). "Different evolutionary processes shaped the mouse and human olfactory receptor gene families." Hum Mol Genet **11**(5): 535-46.

Young, J. M., M. Kambere, et al. (2005). "Divergent V1R repertoires in five species: Amplification in rodents, decimation in primates, and a surprisingly small repertoire in dogs." Genome Res **15**(2): 231-40.

Young, J. M., B. M. Shykind, et al. (2003). "Odorant receptor expressed sequence tags demonstrate olfactory expression of over 400 genes, extensive alternate splicing and unequal expression levels." Genome Biol **4**(11): R71.

Young, J. M. and B. J. Trask (2002). "The sense of smell: genomics of vertebrate odorant receptors." Hum Mol Genet **11**(10): 1153-60.

Zeiske, E., A. Kasumyan, et al. (2003). "Early development of the olfactory organ in sturgeons of the genus Acipenser: a comparative and electron microscopic study." Anat Embryol (Berl) **206**(5): 357-72.

Zhang, X. and S. Firestein (2002). "The olfactory receptor gene superfamily of the mouse." Nat Neurosci **5**(2): 124-33.

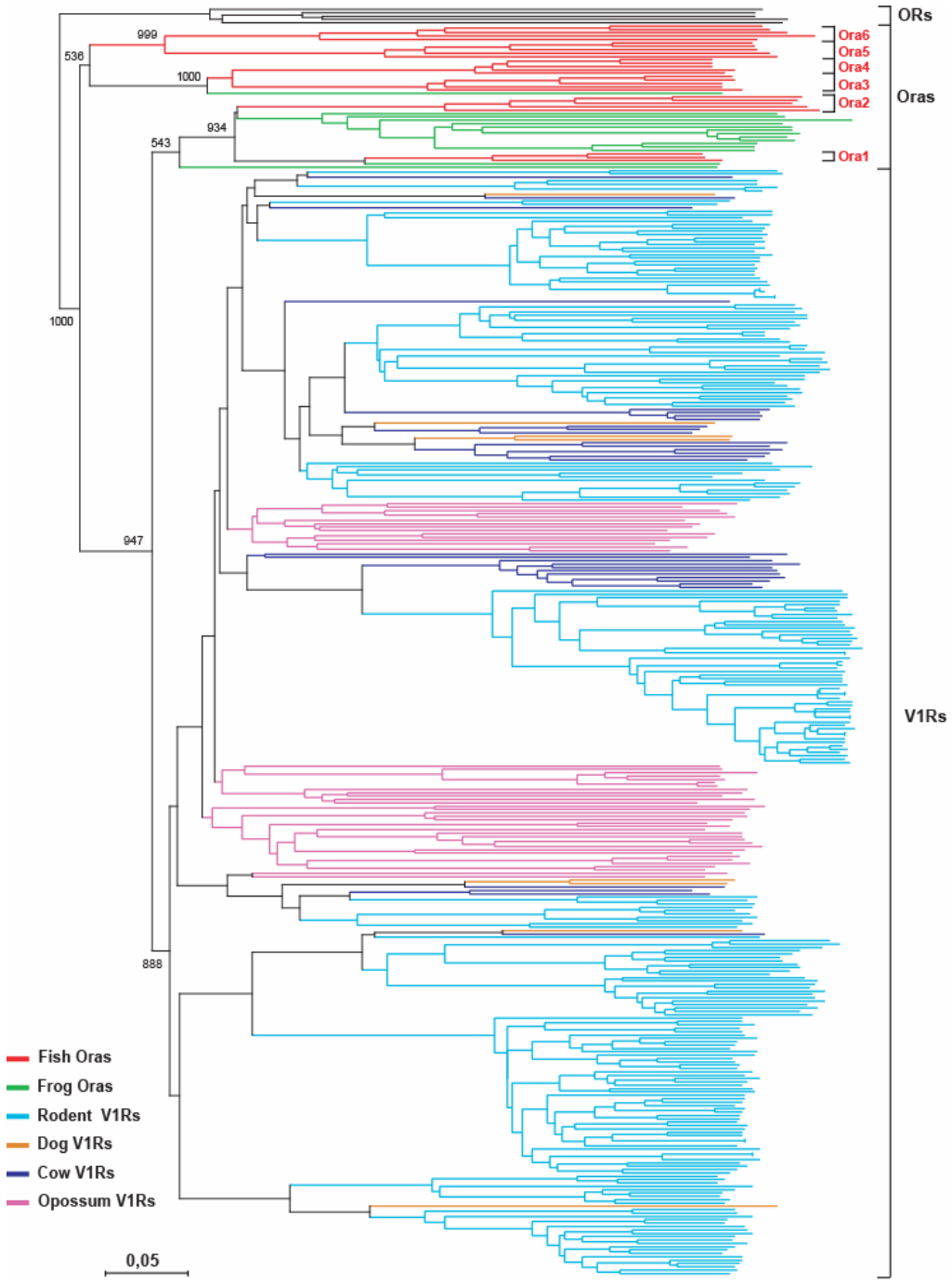
Zhang, X., I. Rodriguez, et al. (2004). "Odorant and Vomeronasal Receptor Genes in Two Mouse Genome Assemblies." Genomics. **83**(5): 802-11.

Zufall, F., S. Firestein, et al. (1994). "Cyclic nucleotide-gated ion channels and sensory transduction in olfactory receptor neurons." Annu Rev Biophys Biomol Struct **23**: 577-607.

XI. APENDIX

A. Abbreviations

A, adenosine
AP, alkaline phosphatase
AOB, accessory olfactory bulb
bp, base pairs
BSA, bovine serum albumine
C, cytosine
cDNA, complementary DNA
DAB, diaminobenzidine
DEPC, diethylpyrocarbonate
DIG, digoxigenin
Dpf, days post fertilization
DNA, desoxynucleic acid
dNTP, desoxynucleotide phosphate
EDTA, ethylenediaminetetraacetic acid
HRP, horseradish peroxidase
IPTG, isopropyl-_-D-1-thiogalactopyranoside
kb, kilo base
LOT, lateral olfactory tract
M, molar
MCS, multiple cloning site
µg, microgram
min, minutes
MOB, main olfactory bulb
MOE, main olfactory epithelium
MOT, medial olfactory tract
ng, nanogram
NGS, normal goat serum
OE, olfactory epithelium
OB, olfactory bulb
OMP, olfactory marker protein
OR, olfactory receptor
OSN, olfactory sensory neuron
PBS, phosphate buffered saline
PCR, polymerase chain reaction
PFA, paraformaldehyde
RNA, ribonucleic acid
SSC, sodium citrate
T, thymidine
TE, tris-EDTA
U, unit
VNO, vomeronasal organ
VR, vomeronasal receptor
VSN, vomeronasal sensory neurons
X-Gal, 5-Bromo-4-chlor-3-indoyl-_-D-galactopyranosid



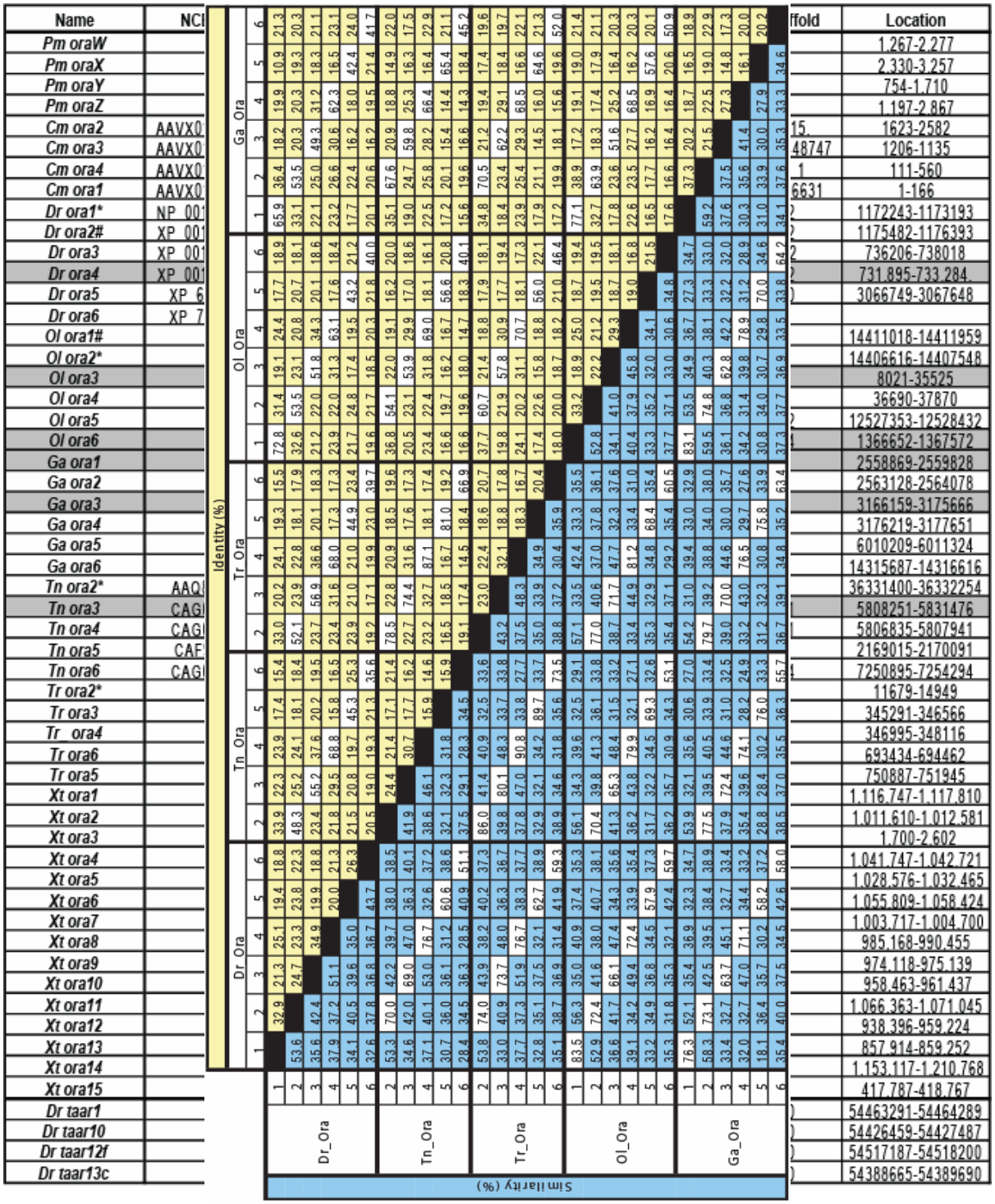
B. Supplemental figures

Figure XI-1 | Phylogenetic tree of all the ora genes plus all the mammalian V1Rs.

Horizontal NJ tree representation, branch length represents phylogenetic distance, scale bar corresponds to 5% replacements. Red lines represent fish *ora* genes, green lines represent frog *ora* genes, light-blue represents rodents (mouse and rat) V1Rs, orange represents dog V1Rs, dark-blue represents cow V1Rs and violet represents opossum V1Rs. A careful analysis of the *Xenopus tropicalis* genome yielded 15 *ora* genes. The somewhat higher number reported very recently (Abbasi, Paparidis et al. 2007) appears to include some duplicate entries. OR genes are used as outgroup. Bootstrap values are indicated at major nodes (1000 trials).

Figure XI-2 | ora and taar genes nomenclature, accession numbers and/or IDs and location.

Gray rows represent manually edited genes; asterisk, see Pfister & Rodriguez, 2005 and double cross, see Shi & Zhang, 2007. The annotation of ora 6 from zebrafish has been discontinued in the NCBI database, but the gene can be retrieved nevertheless. That it exists is shown by the successful cloning from RT-PCR. Only the taar genes where sequences were cloned were included in this table. The remaining ones are as described in (Hussain, Saraiva et al. 2008)



Yellow shaded cells contain values for identity and blue shaded cells contain values for similarity in pairwise comparisons of all the fish *ora* genes. Non shaded boxes indicate the comparisons between orthologs. The main diagonal contains values for paralog comparisons.

Figure XI-3 | Identity and similarity matrix of fish ora genes.

XII. CURRICULUM VITAE

Name: Luís Miguel Rodrigues Saraiva

Birth date and place: 10th October 1980, Portugal

Sex: Male

Nationality: Portuguese

Education:

- June 2008: *summa cum laude* PhD in Genetics, from University of Cologne, Germany ;
- July 2004: Diploma (<> M.Sc.) in Biology with an aggregate of 14 (out of 20) from Évora University, Portugal;
- July 1998: Conclusion of High School (Natural Sciences**)** with an aggregate of 17 (out of 20) from Externato Nossa Sra. Fátima, Manteigas, Portugal.

Research Experience:

- September 2004 – June 2008, University of Cologne, Germany: PhD student, Neurobiology Lab of Sigrun Korsching, Institute of Genetics;
- September 2003 – July 2004, Gulbenkian Institute of Science, Oeiras, Portugal: Diploma Thesis (<> M.Sc.), Populations Genetics Lab of Francisco Dionisio;

Grants & Awards:

- 2008, May: Travel grant from Minerva Foundation to participate in the "28th Blankenese Conference**"**, Hamburg, Germany;
- 2008, May – Travel grant from BCF Jobs 2008 to participate in the "International Talents meet Dutch Life Sciences", Amsterdam, Holland;
- 2007, June – Travel grant from McKinsey & Co. to participate in the "McKinsey's Horizons 2.0 Seminar", Lisbon, Portugal;
- September 2004 - 2008: PhD Fellowship from International Graduate School for Genetics and Functional Genomics, University of Cologne, Germany (4 years support).

Publications:

- Hussain, A*; **Saraiva, LR*** & Korsching SI (**2008**). "*Strong positive selection and high intron dynamics in teleost taar genes*". PNAS *in revision.* *both authors contributed equally
- Kraemer, AM; **Saraiva, LR** & Korsching, SI (**2008**). "*Structural and functional diversification in the teleost S100 family of calcium-binding proteins*". BMC Evolutionary Biology, 8:48
- **Saraiva, LR** & Korsching, SI (**2007**). *"A novel olfactory receptor gene family in teleost fish".* Genome Research. 17:1448-57

Köln, den 6. Juli, 2008