

Evolutionary Dynamics of the OR Gene Repertoire in Teleost Fishes: Evidence of an Association with Changes in Olfactory Epithelium Shape

Maxime Policarpo, Katherine Bemis, James C. Tyler, Cushla J. Metcalfe, Patrick Laurenti, Jean-Christophe Sandoz, Sylvie Rétaux, Didier Casane

▶ **To cite this version:**

Maxime Policarpo, Katherine Bemis, James C. Tyler, Cushla J. Metcalfe, Patrick Laurenti, et al.. Evolutionary Dynamics of the OR Gene Repertoire in Teleost Fishes: Evidence of an Association with Changes in Olfactory Epithelium Shape. Molecular Biology and Evolution, 2021, 38 (9), pp.3742-3753. 10.1093/molbev/msab145 . hal-03280678v2

HAL Id: hal-03280678
https://hal.science/hal-03280678v2

Submitted on 5 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Evolutionary Dynamics of the OR Gene Repertoire in Teleost Fishes: Evidence of an Association with Changes in Olfactory Epithelium Shape

Maxime Policarpo,[1] Katherine E. Bemis,[2] James C. Tyler,[3] Cushla J. Metcalfe,[4] Patrick Laurenti,[5] Jean-Christophe Sandoz,[1] Sylvie Rétaux,[6] and Didier Casane[1,7,*]

[1]Université Paris-Saclay, CNRS, IRD, UMR Évolution, Génomes, Comportement et Écologie, Gif-sur-Yvette, France
[2]NOAA National Systematics Laboratory, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA
[3]Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA
[4]Independent Researcher, Nambour, QLD, PO Box 21, Australia
[5]Université de Paris, Laboratoire Interdisciplinaire des Energies de Demain, Paris, France
[6]Université Paris-Saclay, CNRS, Institut des Neurosciences Paris-Saclay, Gif-sur-Yvette, France
[7]Université de Paris, UFR Sciences du Vivant, Paris, France

***Corresponding author:** E-mail: didier.casane@egce.cnrs-gif.fr.

Associate Editor: David Irwin

Abstract

Teleost fishes perceive their environment through a range of sensory modalities, among which olfaction often plays an important role. Richness of the olfactory repertoire depends on the diversity of receptors coded by homologous genes classified into four families: OR, TAAR, V1R1, and V2R. Herein, we focus on the OR gene repertoire. While independent large contractions of the OR gene repertoire associated with ecological transitions have been found in mammals, little is known about the diversity of the OR gene repertoire and its evolution in teleost fishes, a group that includes more than 34,000 living species. We analyzed genomes of 163 species representing diversity in this large group. We found a large range of variation in the number of functional OR genes, from 15 in the Broad-nose Pipefish *Syngnathus typhle* and the Ocean Sunfish *Mola mola*, to 429 in the Zig-zag Eel *Mastacembelus armatus*. The number of OR genes was higher in species when a multilamellar olfactory rosette was present. Moreover, the number of lamellae was correlated with the richness of the OR gene repertoire. While a slow and balanced birth-and-death process generally drives the evolution of the OR gene repertoire, we inferred several episodes of high rates of gene loss, sometimes followed by large gains in the number of OR genes. These gains coincide with morphological changes of the olfactory organ and suggest a strong functional association between changes in the morphology and the evolution of the OR gene repertoire.

Key words: **OR genes, olfactory rosette, teleosts, gene family dynamics.**

Article

Introduction

Olfaction is an important sensory modality in many animals because it serves essential functions such as feeding, reproduction, migration, kin recognition, and predator avoidance. In vertebrates, odorant molecules are primarily detected by olfactory sensory neurons, each expressing an olfactory receptor, following the one neuron—one receptor rule, although some exceptions have been reported (Sato et al. 2007). Olfactory receptors are G protein-coupled receptors coded by a large gene family (Buck and Axel 1991). The diversity of odors that an individual can discriminate depends on the richness of its olfactory gene repertoire, which results from the number and diversity of functional genes present in its genome. There is a great range of variation in the number and diversity of olfactory receptor genes among species (Nei et al. 2008; Niimura 2009; Khan et al. 2015; Sharma et al. 2019), which may reflect differences in the importance of olfaction relative to other sensory systems. Expansion and contraction of the olfactory receptor gene repertoire in a lineage are governed by a birth-and-death process; that is, recurrent gene duplications and losses that are involved in both the adaptive and nonadaptive evolution of olfaction (Nei et al. 2008).

Vertebrate olfactory receptors genes are coded by four large multigene families named olfactory receptor (OR), trace amine-associated receptor (TAAR), vomeronasal receptor type 1 and 2 (V1R and V2R) (Nei et al. 2008). OR and TAAR genes share sequence similarity and have no introns in the coding sequence. Studies of the evolution of the

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

Downloaded from https://academic.oup.com/mbe/article/38/9/3742/6265480 by guest on 05 November 2021

vertebrate olfactory gene repertoire have focused mainly on mammalian OR genes, and far less research has been undertaken on the OR genes of bony fishes (Osteichthyes). The average number of functional OR genes is about 800 in mammals, but the range of variation is very large, from 58 in the Common Bottlenose Dolphin *Tursiops truncatus* to 2,514 in the African Elephant *Loxodonta africana* (Hughes et al. 2018). In teleosts, the average number of OR genes is an order of magnitude smaller. Although few species of teleosts have been analyzed in prior studies, a wide range of variation has been observed (Niimura 2009; Jiang et al. 2019).

Vertebrate OR genes are classified into two types and eleven families that diverged before the separation of actinopterygians and sarcopterygians: Type 1 (α, β, γ, δ, ε, and ζ) and Type 2 (η, $\theta 1$, $\theta 2$, κ and λ). There is a clear separation between Type 1 and Type 2 genes, but the phylogenetic relationships among the gene families belonging to each type are not well established (Niimura and Nei 2005; Niimura 2009). In tetrapods, the γ family is greatly expanded and, to a lesser extent, the α family also is expanded. Sizes of the other gene families decreased and some disappeared in some lineages. In amniotes, most OR genes belong to the α or γ families. In contrast, teleosts lack α genes and only one γ gene exists in some species, but all other gene families are present (Lv et al. 2019). These observations led to the hypothesis that α and γ genes serve to detect airborne odorants while other gene families detect water-soluble odorants (Niimura and Nei 2005). Hence, gene families would have expanded or contracted differentially after the transition from an ancestral aquatic and aerial olfactory system, in the last common ancestor of bony fishes, to either only aerial olfaction in tetrapods or only aquatic olfaction in teleosts. At smaller phylogenetic scales, high rates of OR gene death have been linked to major ecological transitions. For example, cetaceans lost many OR genes because of their return to the sea (Liu et al. 2019). In primates, an acceleration of OR gene loss that occurred in the ancestral branch of haplorhines is associated with the acquisition of acute vision, and a high rate of OR gene loss was also observed at the ancestral branch of leaf-eating colobines related to the dietary transition from frugivory to folivory (Niimura et al. 2018). Conversely, an extreme expansion of the OR gene repertoire occurred in elephants, although the reason for this is unknown (Niimura et al. 2014). A recent study analyzed how the OR gene repertoire has undergone expansion and contraction across Mammalia with respect to several ecological adaptations (Hughes et al. 2018). Another study found a correlation between the number of OR genes and the relative size of the cribriform plate, a part of the ethmoid bone, and the area of its foramina, which are indicators of the relative number of olfactory sensory neurons (Bird et al. 2018).

In contrast, we know little about the evolution of OR in teleosts. With >34,000 species, teleosts represent about half of the extant vertebrate species and more than 95% of aquatic vertebrates and are an ideal model to study the evolution of the olfactory system in aquatic environments. Although fossils clearly assignable to Teleostei do not appear until the Mesozoic (252 Ma), the compilation of 24 molecular datings (Kumar et al. 2017) suggest that the last common ancestor of extant teleosts may have lived earlier, 284 Ma (244–294 Ma). Regardless of the exact age of teleosts, a vast period of time allowed great diversification of morphology, physiology, behavior, ecology, and habitat. Switches in sensory modalities have involved large numbers of gene births and deaths in teleosts. For example, several deep-sea fishes independently evolved rod opsin-based dim light vision relying on the expansion of RH1 gene repertoires (Musilova et al. 2019). In some cavefishes living in total darkness, loss of vision allowed the decay of many eye-specific genes (Policarpo et al. 2021). Much less is known about the evolution of olfaction and the dynamics of the OR gene repertoire size.

Since the seminal work of Burne at the beginning of the 20th century (Burne 1909), the relative importance of olfaction among sensory systems in teleosts has been assessed by studying the shape and organization of the olfactory epithelium. Most teleosts have a multilamellar olfactory epithelium that forms a rosette, with each lamella increasing the epithelial surface, suggesting that the olfactory system is well developed (Kasumyan 2004). The presence of a rosette is the ancestral state for teleosts (Hansen and Zielinski 2005), indicating that the sense of smell was well developed in their last common ancestor. However, simpler organizations have been observed in many species. For example, the olfactory epithelium is flat (0 lamellae) in the Broad-nose Pipefish *Syngnathus typhle* (Dymek et al. 2021), and there is only one lamella in the Round Goby *Neogobius melanostomus* (Hansen and Zielinski 2005). When a multilamellar olfactory rosette is present, the number of lamellae is highly variable, with up to 230 in the Barred Pargo *Hoplopagrus guentherii* (Hara 1975). Species with a simple olfactory epithelium (a flat olfactory epithelium or an olfactory rosette with very few lamellae), such as the Three-spined Stickleback *Gasterosteus aculeatus*, Crowned Seahorse *Hippocampus coronatus*, Green Pufferfish *Dichotomyctere fluviatilis* are assumed to have degenerate olfaction. Conversely, species that possess olfactory rosettes with many lamellae, such as the Japanese Eel *Anguilla japonica*, Zig-zag Eel *Mastacembelus armatus*, Channel Catfish *Ictalurus punctatus*, and Catla *Labeo catla*, may rely primarily on olfaction to perceive features of their environment. Although many teleost genomes are now available, correlation between the complexity of the olfactory rosette and the richness of the OR gene repertoire has not been investigated. To date, OR gene repertoires of teleosts have been reported for only a handful of species (Niimura and Nei 2005; Niimura 2009; Gao et al. 2017; Jiang et al. 2019; Lv et al. 2019), and no comparative analyses have been made.

The purpose of our study was to describe the evolutionary dynamics of the OR gene repertoire across the diversity of teleosts. We conducted large-scale phylogenetic analyses of the OR gene repertoires of 163 species to examine correlations between contractions and expansions of the OR gene family and changes in morphological traits, ecological parameters, and genome size. We found a positive correlation between the richness of the OR gene repertoire and both the presence of an olfactory rosette and the number of lamellae. In particular, we discovered that the two highest death rates

Downloaded from https://academic.oup.com/mbe/article/38/9/3742/6265480 by guest on 05 November 2021

of OR genes coincide with two independent transitions from a multilamellar olfactory rosette to a simple olfactory epithelium. Moreover, a reacquisition of a multilamellar olfactory rosette coincides with a re-expansion of the OR repertoire. These observations suggest a strong functional link between changes in the morphology of the olfactory organ and the evolution of the OR gene repertoire.

Results

Diversity and evolution of OR gene repertoire in teleost fishes

The genomes of 307 teleost species and two nonteleostean actinopterygians (the Sterlet Sturgeon *Acipenser ruthenus* and the Spotted Gar *Lepisosteus oculatus*) were downloaded from NCBI and ENSEMBL databases.

We discarded six genomes for which the genome assembly size was not similar to the genome size estimated with other methods and compiled in the Animal Genome Size Database (Gregory 2020) (fig. 1*A* and supplementary fig. S1, Supplementary Material online). The completeness of the remaining 301 genomes was assessed using BUSCO (Waterhouse et al. 2018). We discarded all genomes for which <90% of the BUSCO genes were retrieved as complete coding sequences. This threshold allowed us to select a priori 163 teleost and two nonteleost genomes for which we could expect to identify a nearly complete set of OR genes. A higher threshold would have removed more species from consideration and would not have greatly improved the average quality of the genomes analyzed (fig. 1*B*).

We found 27,320 OR genes in the genomes of these 165 species. Among the OR genes, 19,248 were complete with seven transmembrane domains, and we refer to these as functional olfactory receptors; 4,868 genes carried at least one frameshift or a stop codon (also called loss-of-function [LoF] mutations) and were classified as pseudogenes. We also identified 2,344 truncated genes, which were incomplete sequences without any LoF mutations, and 860 edge genes, which were incomplete coding sequences located at scaffold edges (fig. 2*A*).

The number of truncated genes per genome was correlated with the number of pseudogenes ($r = 0.62$, P value = 2.2e−16), whereas the number of edge genes was not correlated with the number of pseudogenes ($r = 0.045$, P value = 0.57), or with the number of truncated genes ($r = 0.036$, P value = 0.65) (respectively supplementary fig. S2*A*–*C*, Supplementary Material online). These results suggest that a large fraction of truncated genes are due to large deletions and correspond to actual pseudogenes, whereas edge genes are mainly assembly artifacts.

We found a positive correlation between the number of pseudogenes and the number of functional genes ($r = 0.42$, P value = 2.5e−08) (supplementary fig. S3*A*, Supplementary Material online). The phylogenetic signal (i.e., the tendency of closely related species to resemble each other more than species drawn at random from the same phylogenetic tree) for both the number of functional ORs and the number of pseudogenes was strong (Pagel's $\lambda = 0.77$ and 0.94, respectively, and both significantly different from 0). Therefore, we applied a phylogenetic correction, after which the correlation still holds true ($r = 0.36$, P value = 2.1e−6) (supplementary fig. S3*B*, Supplementary Material online). This suggests that OR repertoire remodeling is a slow process. Conversely, there was no correlation between the proportion of pseudogenes and the number of functional genes ($r = -0.036$, $P = 0.65$) (supplementary fig. S2*D*, Supplementary Material online), suggesting that the rate of gene loss does not increase with the number of functional genes. These results are similar to those obtained with the OR genes of placental mammals (Niimura et al. 2014).

We assessed the relative efficacy of our pipeline to identify the OR gene repertoire in a genome, including both functional and nonfunctional sequences, by comparing the number of genes retrieved in the present study and in previous analyses of the OR repertoire in teleosts, that is the Zebrafish *Danio rerio*, Three-spined Stickleback *Gasterosteus aculeatus*, Medaka *Oryzias latipes*, Tiger Pufferfish *Takifugu rubripes*, Green Pufferfish *Dichotomyctere nigroviridis* (Niimura 2009), Channel Catfish *Ictalurus punctatus* (Gao et al. 2017), Chinese Perch *Siniperca chuatsi* (Lv et al. 2019), and Mariana Snailfish *Pseudoliparis swirei* (Jiang et al. 2019). We systematically found more OR genes than in these previous reports, and, for several species, many more functional sequences (table 1). This is probably because of the higher quality of more recent genome versions when compared with older versions used in previous studies. This may also result from a better pipeline implementation compared to other analyses based on the same genome versions (more details are given in supplementary fig. S4*A*–*H*, Supplementary Material online).

OR genes were classified into seven previously defined families (Niimura and Nei 2005): α, β, γ, ζ, ε, δ (Type 1) and η (Type 2). Genes of the α family are absent in teleosts, but they are present in the two nonteleost outgroups, *Acipenser ruthenus* and *Lepisosteus oculatus* (fig. 2*A*). Type 2 genes belonging to the families θ, κ, and λ were not analyzed because they may not be OR genes, most often present as a unique copy in vertebrates when they are not absent, with very rare cases of gene gains and losses (Niimura 2009).

To visualize gene sequence diversity within each gene family, we reconstructed a phylogenetic tree with the functional OR sequences of 38 species belonging to 38 teleost orders, each species having the most complete genome within that order. At the scale of teleosts, this phylogeny (supplementary fig. S5, Supplementary Material online) shows that the δ family contains many more genes and is much more diverse than the other families, followed by the ζ and η families. The β and ε families are much less diverse, and diversity is very low in the γ family. This observation also holds at the species level (fig. 2*A*).

Contrasting Richness of the OR Repertoire among Teleost Fishes

We found an unanticipated large range of variation in the number of OR genes among teleosts (fig. 2, supplementary fig. S6, Supplementary Material online). The mean number of

Downloaded from https://academic.oup.com/mbe/article/38/9/3742/6265480 by guest on 05 November 2021

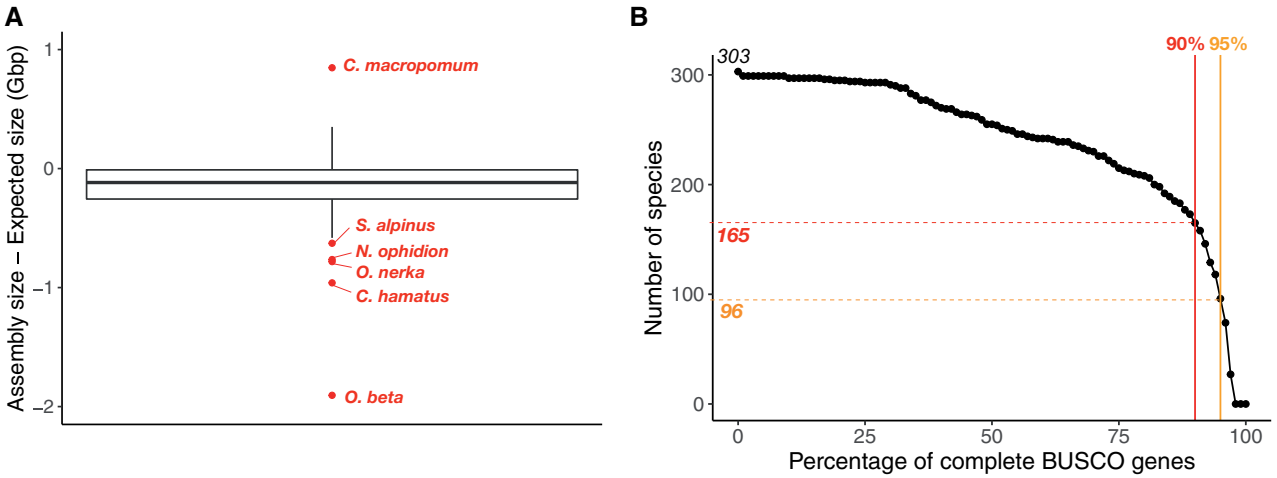


FIG. 1. Genome selection. (*A*) Congruence of genome assembly size and expected genome size based on haploid DNA content (C-value × 0.978 × 10^9) for 309 actinopterygian fishes (307 teleosts and 2 nonteleosts); six incongruent estimates of genome size are shown in red and species names indicated (more details in supplementary fig. S3, Supplementary Material online). (*B*) Number of genomes according to a completeness threshold (% of complete BUSCO genes). Out of 303 genomes with congruent estimates of genome size, 165 (163 teleosts and 2 nonteleosts) show a completeness > 90%, and 96 genomes (96 teleosts) a completeness > 95%.

functional OR genes per species was 117, with the smallest number (15) in the Broad-nose Pipefish *Syngnathus typhle* (Syngnathiformes) and the Ocean Sunfish *Mola mola* (Tetraodontiformes) and the highest (429) in the Zig-zag Eel *Mastacembelus armatus* (Synbranchiformes). The other Syngnathiformes species studied (*Syngnathus acus* and *Hippocampus comes*) were also OR-poor, and only a few δ and η genes were present (fig. 2*B*). In Tetraodontiformes, including *Mola mola*, *Thamnaconus septentrionalis*, *Dichotomyctere nigroviridis*, and three *Takifugu* species, the number of OR genes was more variable (fig. 2*C*).

The number of gene gains and losses along each branch of the phylogenetic tree was inferred using the gene tree—species tree reconciliation method (Chen et al. 2000), with one gene tree per OR gene family (fig. 2*A*). Birth and death rates were computed using these inferred numbers of gene gains and losses (supplementary fig. S7, Supplementary Material online). The mean birth and death rates were very similar, 0.0071 and 0.0073 per gene per million years, respectively (fig. 3), but the death rate was particularly high in the branches leading to Tetraodontiformes (0.3) and Syngnathiformes (0.32).

We also found two cases of high birth rates. The highest birth rate (0.2) was at the level of a deep internal node, in the common ancestor of Labriformes and Cyprinodontiformes while the second-highest birth rate (0.15) was observed in the common ancestor of *Perca* + *Sander* (fig. 2, supplementary fig. S7, Supplementary Material online).

The number of gene losses inferred in terminal branches correlated with the sum of the number of pseudogenes and truncated genes (fig. 4). Interestingly, the correlation was higher when considering only the shortest terminal branches ($r > 0.7$ with maximum branch length < 7 Myr) and decreased progressively when longer terminal branches were added ($r = 0.17$ with all terminal branches, see also supplementary fig. S8, Supplementary Material online). Similar results were obtained using either the number of pseudogenes or the number of truncated genes, but no correlation was found with edge genes (supplementary fig. S9, Supplementary Material online).

We ran simulations of sequence evolution involving nucleotide substitutions and small indels taking into account their relative frequencies estimated in a study of gene decay in teleosts (Policarpo et al. 2021). We found that about half of the pseudogenes were not detected after ~ 40 million generations (supplementary fig. S10, Supplementary Material online). If we also consider that large deletions can remove large fragments of nonfunctional genes and even complete sequences, then the time required for a large fraction of pseudogenes to fade away is probably much shorter. The continuous process of divergence and removal of nonfunctional genes may explain why the expected correlation between the inferred number of gene losses and the observed number of pseudogenes and truncated genes is high only when short terminal branches are analyzed.

Evolution of Olfactory Epithelium Folding in Teleost Fishes

Using the most recent hypotheses about teleost phylogeny and a compilation of descriptions on the olfactory organ, we inferred the evolution of the olfactory epithelium shape using maximum likelihood and parsimony methods. We conducted a literature survey of the shape of the olfactory epithelium for 220 actinopterygians beginning with the work of Burne (1909) (supplementary Data S1, Supplementary Material online). Our analysis indicated that a multilamellar olfactory rosette is ancestral, which was expected (Hansen and Zielinski 2005), but we also refined aspects of its evolution. In particular, the multilamellar olfactory rosette has been lost several times and regained in some lineages (fig. 2,

Downloaded from https://academic.oup.com/mbe/article/38/9/3742/6265480 by guest on 05 November 2021

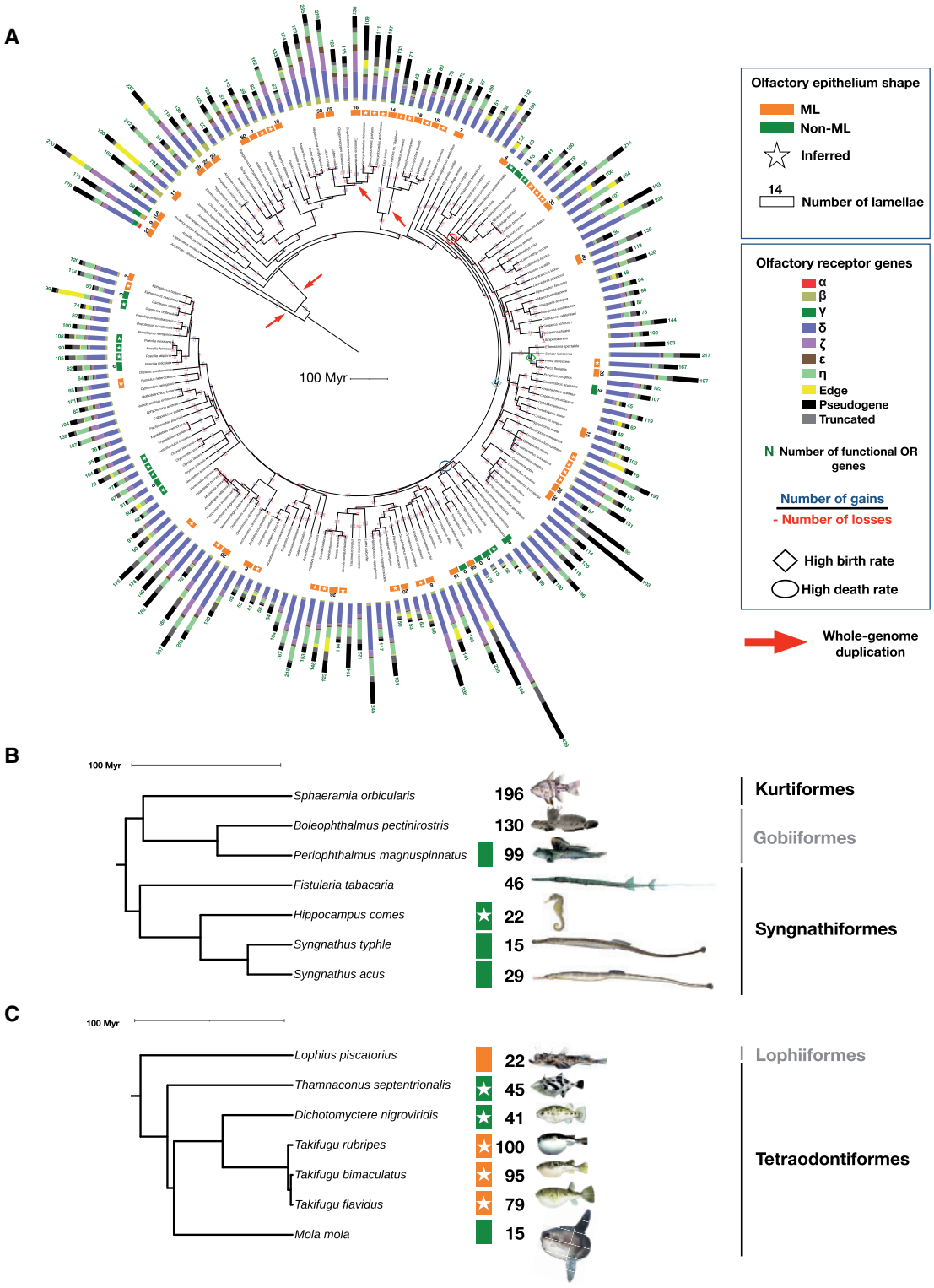


FIG. 2. Chronograms. (*A*) Dated phylogeny for 163 teleost fishes and two outgroups (from https://fishtreeoflife.org/). For each species, the total number of functional OR genes, the proportion of genes from each gene family, the proportion of truncated and edge genes, and the proportion of pseudogenes are provided. When available from the literature, presence or absence of a multilamellar olfactory rosette is reported (ML/non-ML). If found in the literature, the number of olfactory lamellae is also indicated. Presence or absence of a multilamellar olfactory rosette was inferred for some species without data on the shape of the olfactory epithelium using information available for a congeneric species (white star); the number of lamellae was not inferred because of intrageneric variation. Inferred number of gene gains and losses are provided on each branch (corresponding birth and death rates are given in supplementary fig. 7, Supplementary Material online). The branches associated with the two highest birth rates and the two highest death rates are indicated by diamond and oval symbols, respectively. Whole-genome duplications are indicated by red arrows. The tree was annotated and visualized using iTOL. (*B*) Focus on Syngnathiformes and related species. (*C*) Focus on Tetraodontiformes and related species.

Downloaded from https://academic.oup.com/mbe/article/38/9/3742/6265480 by guest on 05 November 2021

Table 1. Comparison of the number of OR genes retrieved from the genome of eight species of teleosts in the present and previous studies.

	Present study			Previous studies			References
	F	I	P	F	I	P	
Danio rerio	162	4	8	154	1	21	Niimura (2009)
Gasterosteus aculeatus	107	11	33	102	5	52	Niimura (2009)
Oryzias latipes	81	3	5	68	6	24	Niimura (2009)
Takifugu rubripes	100	17	12	47	39	39	Niimura (2009)
Dichotomyctere nigroviridis	41	17	18	11	4	19	Niimura (2009)
Siniperca chuatsi	144	12	43	123	N/A	29	Lv et al. (2019)
Ictalurus punctatus	87	11	14	27	N/A	20	Gao et al. (2017)
Pseudoliparis swirei	48	9	10	43	0	10	Jiang et al. (2019)

NOTE.—F, functional genes; I, incomplete genes (truncated and edge); P, pseudogenes; N/A, data not available.
More details on this analysis can be found in supplementary fig. S4, Supplementary Material online.

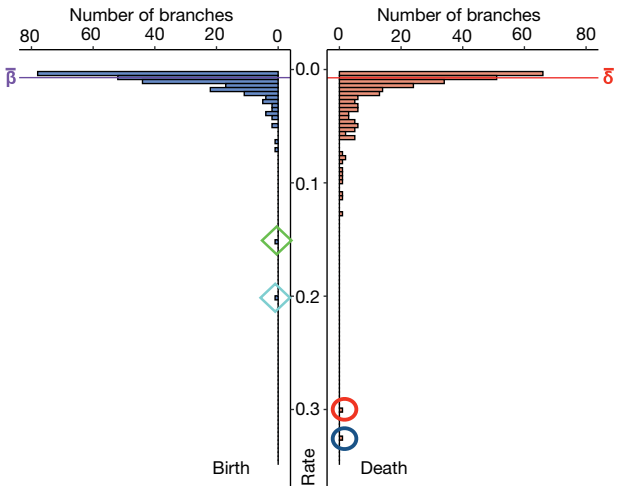


FIG. 3. Distribution of birth and death rates along the branches of the phylogenetic tree shown in fig. 2A. The mean values are indicated and the two highest birth rates and the two highest death rates are highlighted with diamond and oval frames, respectively.

supplementary fig. S11, Supplementary Material online). Maximum likelihood and parsimony methods gave similar results, except for a few internal nodes, such as the most recent common ancestor of Syngnathiformes and Gobiiformes + Kurtiformes, which is inferred to have a multilamellar olfactory rosette with the maximum likelihood MPPA and parsimony with DELTRAN (supplementary fig. S11*A* and *C*, Supplementary Material online), undefined with DOWNPASS (supplementary fig. S11*B*, Supplementary Material online) but a flat olfactory epithelium was inferred with ACCTRAN (supplementary fig. S11*D*, Supplementary Material online).

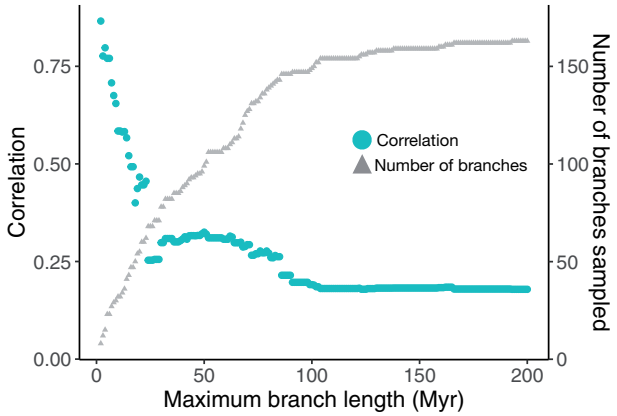


FIG. 4. Variation in the correlation between the number of gene losses inferred using a gene tree—species tree reconciliation method and the number of observed pseudogenes and truncated genes in the external branches of the phylogenetic tree according to a maximum branch length threshold ranging from 2 to 200 Myr. For each threshold, the number of branches is indicated by a gray triangle.

Correlation of the size of the OR repertoire with morphological and ecological traits

We examined correlations between olfactory epithelium shape, the number of olfactory lamellae, and the size of the OR gene repertoire. There is enormous diversity in the anatomy of olfactory epithelia, but herein we distill this information to a binary character for our analysis: (1) multilamellar olfactory rosette present (3 or more primary olfactory lamellae); (0) Multilamellar olfactory rosette absent (0–2 primary olfactory lamellae).

Among the 220 species for which we had information about olfactory epithelium shape, we had estimates of the size of the OR gene repertoire for 41 species. For 37 of the 41 species, we found information about the mean number of primary olfactory lamellae. For 38 additional species for which we had an estimate of the size of the OR gene repertoire, we inferred olfactory epithelium shape using information available for at least one species in the same genus (fig. 2, supplementary Data S1, Supplementary Material online), assuming that closely related species have similar olfactory epithelium (presence or absence of multilamellar olfactory rosette). However, we did not use information about the number of lamellae in a given species to infer this character at the genus level because it is more labile (Kasumyan 2004).

Species without a multilamellar olfactory rosette had significantly fewer functional OR genes than species with a multilamellar olfactory rosette (fig. 5*A* and *B*). Olfactory epithelium shape had a strong phylogenetic signal (Pagel's $\lambda = 0.999$, significantly different from 0). To account for the phylogenetic signal, we used a phylogenetic logistic model; the correlation between the olfactory epithelium shape and the number of OR genes was still significant (fig. 5*A* and *B*, supplementary fig. S12, Supplementary Material online). This correlation was statistically significant with the whole data set (*P* value = 0.035, 79 species, including the 38 species for which we inferred olfactory epithelium shape, fig. 5*A*) and with the reduced data set containing only 41 species for which

Downloaded from https://academic.oup.com/mbe/article/38/9/3742/6265480 by guest on 05 November 2021

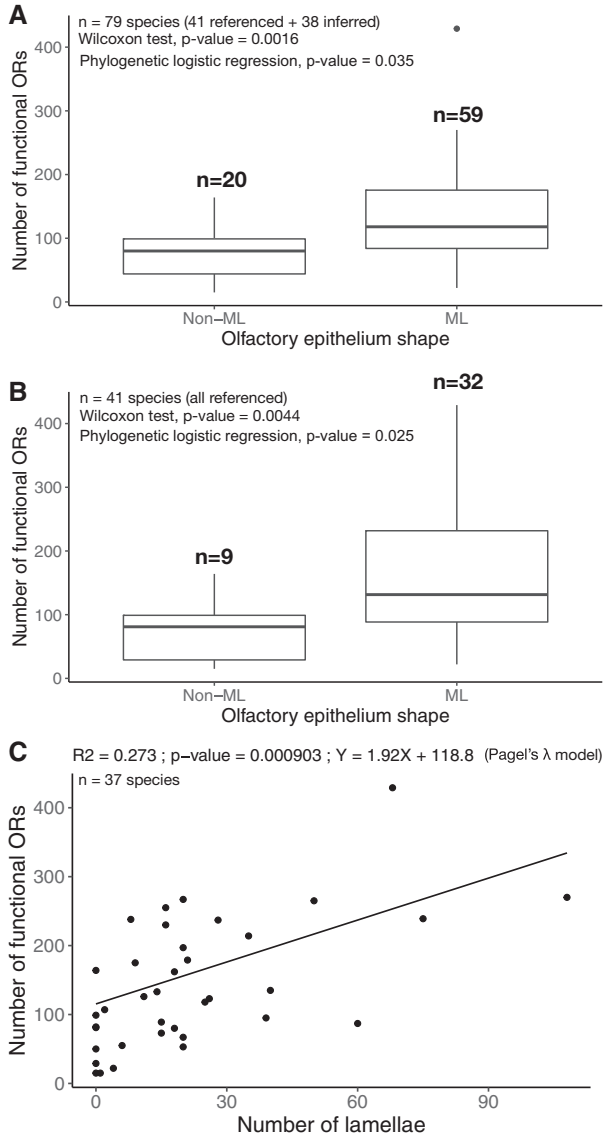


FIG. 5. Correlation between the number of functional OR genes and olfactory epithelium characteristics. (*A*) Correlation with absence (non-ML) or presence (ML) of a multilamellar olfactory epithelium, without taking into account the phylogeny (Wilcoxon test) or taking into account the phylogeny (phylogenetic logistic regression) and considering species for which data on the olfactory epithelium are either available or are available for a congeneric species. (*B*) Same analysis as (*A*), but restricted to species for which data are available. (*C*) Correlation between the number of lamellae and the number of functional OR genes taking into account the phylogeny (phylogenetic linear regression with a Pagel's λ model).

olfactory epithelium shape was directly known from the literature (P value = 0.025, fig. 5*B*).

We also found a positive correlation between the number of lamellae and the number of functional OR genes ($r = 0.58$, P value = 0.00015), which was also significant after phylogenetic correction with Pagel's λ model ($r = 0.52$, P value = 0.0009, AIC = 432.8) or an Ornstein-Uhlenbeck model ($r = 0.5$, P value = 0.0018, AIC = 431.2), and a Brownian motion model ($r = 0.33$, P value = 0.0432, AIC = 435.9) (fig. 5*C*). However, there was no correlation between the number of pseudogenes and olfactory epithelium shape, or the number of lamellae (supplementary fig. S13, Supplementary Material online).

We tested for a correlation between the number of functional OR genes and other morphological characters (maximum length of the fish, relative eye size (eye diameter/standard length of the fish), ecological parameters (trophic level, preferred temperature, and maximum depth) which were retrieved from FishBase (https://www.fishbase.us), and genome size. No significant correlations were found (supplementary fig. S14, Supplementary Material online).

Discussion

We examined the evolutionary dynamics of the size of the OR gene repertoire in a broad sample of teleost fishes and found a larger range of variation than previously recognized (Niimura 2009; Zhu et al. 2017; Jiang et al. 2019; Lv et al. 2019). Large repertoires do not seem to result from bursts of gene duplication, but we identified several episodic high rates of gene pseudogenization that occurred at the root of clades with species that have very few OR genes. Our analysis indicates that the richness of the OR gene repertoire correlates with the presence of a multilamellar olfactory rosette. We found that some large gene losses coincide with a simplification of the olfactory epithelium and that several cases of secondary complexifications of the olfactory epithelium coincide with secondary expansions of the OR gene repertoire.

OR Functions in Teleost Fishes

In the Zebrafish *Danio rerio*, five types of olfactory sensory neurons (ciliated, microvillus, crypt, kappe, and pear) with distinct morphologies, connections, and functions, are found in the olfactory epithelium. Segregated neural pathways are responsible for coding and processing information from different types of olfactory stimuli, at least at the level of the olfactory bulb (Friedrich and Korsching 1997; Biechl et al. 2017; Calvo-Ochoa and Byrd-Jacobs 2019). Less is known about molecular components and functional organization of the olfactory system in other teleosts, but similar studies in the Channel Catfish *Ictalurus punctatus* (Hansen et al. 2003) and the Goldfish *Carassius auratus* (Hansen et al. 2004) indicate that they may be largely conserved in teleosts. OR genes are expressed in ciliated neurons that recognize bile acids and project to medial glomeruli in the olfactory bulb (Yoshihara 2014). Bile acids are biliary steroids synthesized in the liver, secreted in the intestine and released to the environment. These molecules are used as pheromones in the context of social and reproductive behaviors and convey information about the status of the emitter (Doyle and Meeks 2018). For example, one of the most studied and common bile acids, taurocholic acid, acts as an attractant at very low concentration (10^{-8} M), to which females of *D. rerio* are more sensitive to than males (Michel and Lubomudrov 1995).

We know little about the molecules recognized by the different ORs in teleost fishes because no OR has been deorphanized so far. Although current evidence suggests that OR genes may be expressed in sensory neurons involved in social behaviors (e.g., reproduction, migration, kin

Downloaded from https://academic.oup.com/mbe/article/38/9/3742/6265480 by guest on 05 November 2021

recognition, and social status) rather than foraging and feeding, more functional data are needed to describe the odorant response range of teleost ORs.

Diversity and evolution of the OR gene repertoire in teleost fishes

We used the many genomes now available to identify OR gene repertoires in 163 high-quality genomes from species representative of teleost diversity. We evaluated the accuracy of our OR gene mining pipeline by comparing our results with similar analyses performed previously on eight species. In comparison to prior studies, we systematically recovered more OR genes, sometimes finding that the initial estimate greatly underestimated the number of OR genes. We found that some species previously thought to have few functional OR genes, such as *Dichotomyctere nigroviridis* (11 genes) and *Ictalurus punctatus* (27 genes) possess larger OR gene repertoires (41 and 87, respectively) and are not among the fishes with the lowest number of OR genes.

We found an unanticipated large range of variation of OR gene number among teleosts, with an average number of 117 OR genes. The lowest number of OR genes (15) was observed in two distantly related species, the Broad-nose Pipefish *Syngnathus typhle* and the Ocean Sunfish *Mola mola*. In both cases, the few remaining intact ORs primarily belong to the δ and η gene families, which are among the three largest families in teleosts. In both *S. typhle* and *M. mola*, very few pseudogenes were identified, suggesting that contraction of the OR gene repertoire is ancient.

The highest number of OR genes (429) was found in the Zig-zag Eel *Mastacembelus armatus* (Synbranchiformes). This is primarily the result of duplications of genes belonging to δ and ζ families, that are among the three largest OR gene families. The high percentage of pseudogenes indicates high gene turnover in this lineage.

Together, these results suggest that in the most extreme cases of reduction and expansion of the OR gene repertoire, the implication of different OR gene families is not strongly biased.

Within teleosts, gene turnover seems to be driven by slow birth-and-death evolutionary dynamics involving similarly low birth and death rates. However, we identified a few instances of high birth and death rates. The highest birth rate was observed at the level of a deep internal branch (fig. 2). Because it was followed by various death rates in more recent branches, it is not correlated with extant species that have a rich OR gene repertoire. The second highest birth rate occurred in the common ancestor of *Sander* + *Perca* spp., a split that occurred relatively recently, and, in this case, is associated with high numbers of OR genes, 167 in the Yellow Perch *Perca flavescens* and 217 in the Zander *Sander lucioperca*, compared to the teleost average of 117. More striking is the link between the two highest death rates and the two species with the smallest OR gene repertoires, indicating ancient massive gene losses without compensation by gene duplications in some lineages.

An extensive study of the dynamics of OR gene repertoires in mammals found a low correlation ($r = 0.31$) between the number of gene losses inferred using the reconciled-tree method and the number of observed pseudogenes (Niimura et al. 2014). Our analyses suggest an explanation for this counter-intuitive result. Indeed, using all terminal branches, we found a similarly low correlation ($r = 0.17$) for teleosts. However, using only data for short terminal branches, which corresponds to the most recent evolutionary events in independent lineages, the correlation was much higher ($r > 0.7$ when using only branches shorter than 7 Myr). This suggests that the oldest pseudogenes are not always identified because they are too divergent or because they are removed from genomes by large deletions, hampering accurate estimation of the number of pseudogenes in long terminal branches. Analyses at smaller phylogenetic scales showed an expected congruence of estimates of the rate of gene losses based on these independent approaches.

Number of OR Genes and Correlation with Olfactory Epithelium Complexity

We searched for correlations between the size of the OR gene repertoire and morphological traits, ecological parameters and genome size. We did not find any correlation with fish maximum length, relative eye size, maximum depth range, trophic level, temperature preference, or genome size. In contrast, the shape of the olfactory epithelium (absence or presence of a multilamellar olfactory rosette) and, when a multilamellar olfactory rosette is present, the number of its lamellae, were positively correlated with the number of OR genes. These correlations suggest that epithelial folding allows for the presence of more sensory neurons and also a more diversified OR repertoire. Moreover, these correlations indicate a relationship between morphological and genomic changes in accordance with the relative importance of olfaction among the suite of sensory systems.

Case Studies

It is instructive to study cases of extreme OR gene repertoire reduction and subsequent evolution at the molecular and morphological levels.

One case concerns the Broad-nosed Pipefish *Syngnathus typhle* and closely related species. The inference of gene gains and losses indicates a large number of gene losses in the common ancestor of the clade Syngnathiformes + (Kurtiformes + Gobiiformes). If the phylogenetic relationship between Syngnathiformes and (Kurtiformes + Gobiiformes) is correct, then the disappearance of the multilamellar olfactory rosette in Syngnathiformes is closely related with a massive reduction of the OR gene repertoire. In the common ancestor of Syngnathiformes, the number of OR genes also greatly declined. The tiny repertoire of OR genes could be, with a flat olfactory epithelium, a shared character of Syngnathidae (seahorses, pipefishes, and seadragons). Behavioral studies suggest that vision is the leading sense during mating, and that olfaction is less important in *S. typhle* (Berglund and Rosenqvist 2001; Lindqvist et al. 2011). In the clade Gobiiformes + Kurtiformes, a re-expansion of the OR gene repertoire could have occurred at different rates in different lineages, but no data are available to relate these OR

Downloaded from https://academic.oup.com/mbe/article/38/9/3742/6265480 by guest on 05 November 2021

gene repertoire expansions with changes in olfactory epithelium shape.

The diversity of olfactory epithelium within Tetraodontiformes presents a potential model to study evolution of the olfactory system and the OR gene repertoires. The last common ancestor of Tetraodontiformes most likely had a multilamellar olfactory rosette that is present in three families, Triacanthodidae, Triacanthidae, and Triodontidae (Tyler 1980; Arcila and Tyler 2017) (see also supplementary fig. S11*E–I*, Supplementary Material online). Repeated losses of the multilamellar olfactory rosette occurred within Tetraodontiformes, in particular in Monacanthidae and Tetraodontidae. In both cases, loss of the multilamellar olfactory rosette is associated with a contraction of the OR gene repertoire to 45 functional OR genes in the monacanthid *Thamnaconus septentrionalis* and 41 functional OR genes in the tetraodontid *Dichotomyctere nigroviridis*. *Takifugu* spp., also tetraodontids, have a multilamellar olfactory rosette, but the series of folds do not have the ancestral rosette shape, suggesting that it might be a secondary re-acquisition. We infer an expansion of the OR gene repertoire in the common ancestor of *Takifugu* because the functional OR repertoire size ranges from 79 in *Takifugu flavidus* to 100 in *Takifugu rubripes*. Finally, the small number of functional OR genes (15) in the Ocean Sunfish *Mola mola* corresponds with their anatomy. *Mola mola* has minute nostrils almost flush with the surface of the skin (Tyler 1980). We examined an adult specimen of *M. mola* (129 cm total length) and found that the olfactory epithelium is flat, except for a small V-shaped primary lamella on the surface below the incurrent nare (KEB unpublished data). The brain of *M. mola* has greatly reduced olfactory nerves as well as reduced olfactory bulbs (Burr 1928). Together this morphological data and the small number of OR genes suggest that olfaction is particularly regressed in this species.

We observed a 2.5-fold difference in OR number between the Zebrafish *Danio rerio* and the related cyprinid *Danionella dracula*. The miniaturized and transparent species belonging to *Danionella* have recently attracted the interest of neuroscientists because of the advantages for functional imaging, and of developmental biologists as a model for understanding the regulation of size and developmental heterochronies and truncations (Schulze et al. 2018; Conway et al. 2021). *Danionella dracula* and *D. translucida* have small, skull-less brains but rich behavioral repertoires, including acoustic communication. A reduction in social olfactory communication associated with OR gene losses could be related or compensated for by the emergence of a sound-producing apparatus and acoustic communications skills in these species. Such sensory trade-offs could also explain the relatively small OR gene repertoire of electric fishes *Electrophorus electricus* and *Paramormyrops kingsleyae* compared to their closest relatives, because their social interactions may primarily involve electric communication (Gracheva and Bagriantsev 2018). However, additional analyses are required to investigate the importance of sensory trade-offs in the evolution of the size of OR gene repertoires in teleosts.

Perspectives

What kinds of ecological transitions drove in parallel the loss or reappearance of a multilamellar olfactory rosette and contraction or re-expansion of the OR gene repertoire in different teleost lineages? How can Sea-Horses and Ocean Sunfish thrive with such poor OR repertoires and why do several species of *Takifugu* have a multilamellar olfactory rosette and a large OR gene repertoire relative to other Tetraodontiformes that have a flat olfactory epithelium and a limited OR gene repertoire? Further studies examining social communication, ecology, life history traits, and sensory physiology are needed to answer these fascinating questions.

New high-quality genome sequences are becoming available at an ever-increasing rate, but we found that anatomical data on the olfactory organ is rare and that refined descriptions of the olfactory epithelium shape and organization are available only for a handful of model species. In particular, it would be very interesting to analyze the evolution of the number, distribution, and density of olfactory sensory neurons in relation with the olfactory epithelium shape and the number of OR genes. Future large-scale comparative analyses should include such data for a large sample of teleost species. At the molecular level, an interesting complementary analysis would concern the TAAR gene repertoire. It will be particularly important to examine if the correlation we observed between OR repertoire size and olfactory epithelium shape also holds for the TAAR repertoire size.

Materials and Methods

Assessment of Genome Completeness and OR Gene Mining

The genomes of 307 teleost species and those of two outgroups (*Acipenser ruthenus* and *Lepisosteus oculatus*) were downloaded from NCBI and ENSEMBL databases on 04/10/2020. The names of the species and the genome assembly versions can be found in supplementary Data S1, Supplementary Material online.

For each species, the genome completeness was assessed with BUSCO 3.1.0 (Waterhouse et al. 2018) using the Actinopterygii odb9 database and the expected genome size extracted from the Animal Genome Size Database (Gregory 2020).

We developed a pipeline (https://github.com/MaximePolicarpo/Olfactory_receptor_genes) to mine olfactory receptor genes in each genome following the procedure of Niimura and Nei (2005). Briefly, tblastn (Ye et al. 2006) was used with a set of query sequences that are teleost OR genes identified by Niimura (2009) and an e-value threshold of 1e−10 to find genomic regions containing OR genes. Nonoverlapping best-hit regions were extracted using samtools (Li et al. 2009). These sequences were extended 1,000 bp upstream and 1,000 bp downstream and open reading frames (ORFs) larger than 750 bp were extracted using EMBOSS getorf (Rice et al. 2000). Only ORFs that best matched to a known olfactory receptor against the UniProt database (The UniProt Consortium 2019) were retained. These ORFs were aligned with the teleost OR gene set identified by Niimura

Downloaded from https://academic.oup.com/mbe/article/38/9/3742/6265480 by guest on 05 November 2021

(2009) and several outgroup sequences (ora1: A0A0R4IM31.1, taar13: XP_021328050.1, adra2a: NP_919345.2, adrb1: AAI62819.1, galr2b: XP_001339169.1, hrh1: XP_009304002.1 and htr1aa: NP_001116793.1) using MAFFT v7.310 (Katoh et al. 2005), and a maximum likelihood tree was computed with IQ-TREE (Nguyen et al. 2015). We only kept genes that clustered with known OR genes by visual inspection of trees and removed 100% identical nucleotide sequences using CD-HIT (Fu et al. 2012). Retained sequences were classified as functional ORs.

This set of functional ORs was then used for a new round of tblastn with a more stringent e-value of 1e−20. For each genome, best-hit regions were extracted and discarded if overlapping with a functional OR gene region. Otherwise, the best tblastn match was extracted and sequences were classified as follows: 1) "edge" if the blast target sequence was less than 30 bp away from the scaffold end; 2) "pseudogene" if the blast target sequence contained at least a stop codon or frameshift; 3) "truncated" if the sequence was incomplete but without internal stop codons or frameshifts.

Functional, Truncated and Edge genes were aligned with the teleost OR gene set identified by Niimura (2009) using MAFFT and classified into OR gene families based on their position on a maximum likelihood tree computed with IQ-TREE. Pseudogenes were classified based on their best blastx match.

Phylogenetic Framework and Morphological/Ecological Trait Assembly

A time-calibrated phylogeny of ray-finned fishes was downloaded from https://fishtreeoflife.org (Rabosky et al. 2018). The tree was pruned using the R package ape (Paradis and Schliep 2019) to keep only the 165 species whose genomes we used in our study. Ecological information on each species was extracted from FishBase (https://www.fishbase.se) using the R package rfishbase (supplementary Data S1, Supplementary Material online).

We gathered information about the olfactory epithelium shape and the number of lamellae per olfactory rosette in 220 actinopterygians from an extensive literature survey (supplementary Data S1, Supplementary Material online). Among these species, genomic data were available for 41 species; for an additional 38 species genomic data were available for at least one species belonging to the same genus. For each species, the olfactory epithelium was classified as multilamellar olfactory rosette present (3 or more lamellae), or multilamellar olfactory rosette absent if the olfactory epithelium was flat (0 lamellae) or only a few lamellae (1 or 2 lamellae) were present. Specimens referenced in the literature were adults. We did not take into account sexual dimorphism or individual age, which have been shown to impact number of olfactory lamellae (Kasumyan 2004; Abrahão et al. 2019).

The time-calibrated "fishtreeoflife" was pruned again to retain 205 of the 220 species for which information about the olfactory epithelium was available (15 species were not in the "fishtreeoflife"). The evolution of the olfactory epithelium shape was inferred by maximum likelihood (MPPA with F81 model) and parsimony (with DOWNPASS, DELTRAN, or ACCTRAN methods) using PastML (Ishikawa et al. 2019). Trees were rooted and visualized using iTOL (Letunic and Bork 2007). Species names were validated using Eschmeyer's Catalog of Fishes (Fricke et al. 2021) (supplementary Data S1, Supplementary Material online).

Small Scale OR Genes Phylogenies

To compare the results of our implementation of a method for identifying OR genes (Niimura 2009) with the results obtained in previous studies, sequences of OR genes from eight teleost species (*Danio rerio*, *Dichotomyctere nigroviridis*, *Oryzias latipes*, *Gasterosteus aculeatus*, *Takifugu rubripes*, *Siniperca chuatsi*, *Pseudoliparis swirei*, and *Ictalurus punctatus*) were retrieved from previous studies (Niimura 2009; Gao et al. 2017; Jiang et al. 2019; Lv et al. 2019). The sequences were aligned with those found in the present study using MAFFT (on amino acids) and a ML tree was computed with IQ-TREE. The optimal model was inferred with ModelFinder, while the robustness of the nodes was evaluated with 1,000 ultrafast bootstraps. Trees were visualized, rooted, and annotated using iTOL.

To obtain an overall picture of the diversity of OR gene families in teleosts, we also inferred a phylogenetic tree of all functional OR sequences from 38 species, each belonging to a different order (species and order names can be found in supplementary Data S2, Supplementary Material online). Protein sequences were aligned using MAFFT and a near-ML tree was computed with FastTree (Price et al. 2010), and the robustness of the nodes was evaluated with 1,000 fast-global bootstraps.

OR Birth and Death Rates

We inferred the number of gene gains and losses along each branch of the species tree using a gene tree—species tree reconciliation method. For each OR gene family, an alignment was performed with MAFFT and a maximum likelihood tree was computed with IQ-TREE. Nodes with low bootstrap values (<90%) were collapsed into polytomies using the R package ape. We then used NOTUNG 2.9.1.5 (Chen et al. 2000) to root and reconcile gene trees with the species tree using the phylogenomics option.

Birth and death rates along each branch as well as the mean birth and death rate were computed using the equations of Niimura et al. (2014). We excluded branches with lengths < 2 million years because differences in gene retrieval and genome qualities had a major impact on inferred birth and death rates (supplementary Data S1, Supplementary Material online).

Phylogenetic Comparative Analyses

The phylogenetic signal (Pagel's λ) of each trait was estimated with the function phylosig in the R package phytools with the option test = TRUE (Revell 2012).

The R package phylolm (Tung Ho and Ané 2014) was used to perform phylogenetic linear regressions between continuous traits using the phylolm function with various evolutionary models, and phylogenetic logistic regression between binary and continuous traits using the phyloglm function.

Downloaded from https://academic.oup.com/mbe/article/38/9/3742/6265480 by guest on 05 November 2021

Correlation between the number of gene losses and the number of pseudogenes in terminal branches

For each terminal branch of the species tree, the number of gene losses was inferred using the gene family trees of intact OR genes and the program NOTUNG 2.9.1.5 (Chen et al. 2000) as described above. We also estimated the numbers of OR pseudogenes (OR sequences with at least one LoF mutations), truncated genes (most likely pseudogenes resulting from large genomic deletions) and edge genes (truncated genes most likely resulting from assembly artifacts).

The correlation (Pearson's r) between the number of gene losses and the number of pseudogenes and truncated genes, together or separately, and the number of edge genes was computed for all terminal branches or using only a subset of branches shorter than a given threshold ranging from 2 Myr to 200 Myr.

To estimate the time necessary for a pseudogene to become undetectable due to the accumulation of substitutions and small indels, we ran simulations in which an intact OR sequence accumulated nucleotide substitutions (at a rate of 10-8 per generation) and indels at a rate of 0.05×10^{-8} per generation as estimated previously (Policarpo et al. 2021). We also took into account the transition/transversion ratio and the relative frequency of indels ranging from 1 and 9 bp in length (more details on the numerical values used for the simulations are given in the Python script available at https://github.com/MaximePolicarpo/Olfactory_receptor_genes).

Each time a new mutation was produced in the evolving sequence, a tblastn analysis was performed using the evolving DNA sequence and the translated original OR sequence. The evolving sequence was classified as unrecognized if the e-value was greater than 10^{-20}, which was the threshold used to mine OR pseudogenes, and the number of generations to get an unrecognized sequence registered. Running this simulation 10,000 times allowed the estimate of the distribution of the number of generations necessary to make a pseudogene undetectable.

Data Availability

All data and results of the analyses performed in this study are available for download at figshare (https://figshare.com/articles/dataset/Supplementary_files_-_ORs_in_teleost/14180090): nucleotide sequences of the olfactory receptor genes in three FASTA format files (functional genes, incomplete genes, and pseudogenes); version of genomes used for this study, assembly sizes, expected genome sizes, BUSCO assessment results, olfactory receptor gene count, olfactory epithelium shape, ecological data, and NOTUNG results in Supplementary_Table.xlsx; BUSCO assessment results in graphic format and phylogenetic trees (Newick format).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Matthieu Haudiquet for his help with the R package phylolm.

References

Abrahão VP, Pastana M, Marinho M. 2019. On a remarkable sexual dimorphic trait in the Characiformes related to the olfactory organ and description of a new miniature species of *Tyttobrycon Géry* (Characiformes: characidae). *PLoS One* 14(12):e0226130.

Arcila D, Tyler JC. 2017. Mass extinction in tetraodontiform fishes linked to the Palaeocene–Eocene thermal maximum. *Proc R Soc B*. 284(1866):20171771.

Berglund A, Rosenqvist G. 2001. Male pipefish prefer ornamented females. *Anim Behav*. 61(2):345–350.

Biechl D, Tietje K, Ryu S, Grothe B, Gerlach G, Wullimann MF. 2017. Identification of accessory olfactory system and medial amygdala in the zebrafish. *Sci Rep*. 7:44295.

Bird DJ, Murphy WJ, Fox-Rosales L, Hamid I, Eagle RA, Van Valkenburgh B. 2018. Olfaction written in bone: cribriform plate size parallels olfactory receptor gene repertoires in Mammalia. *Proc R Soc B*. 285(1874):20180100.

Buck L, Axel R. 1991. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* 65(1):175–187.

Burne RH. 1909. The anatomy of the olfactory organ of teleostean fishes. *Proc Zool Soc Lond*. 2:610–663.

Burr HS. 1928. The central nervous system of *Orthagoriscus mola*. *J Comp Neurol*. 45(1):33–128.

Calvo-Ochoa E, Byrd-Jacobs CA. 2019. The olfactory system of zebrafish as a model for the study of neurotoxicity and injury: implications for neuroplasticity and disease. *Int J Mol Sci*. 20(7):1639.

Chen K, Durand D, Farach-Colton M. 2000. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J Comput Biol*. 7(3-4):429–447.

Conway KW, Kubicek, KM Britz, R. 2021. Extreme evolutionary shifts in developmental timing establish the miniature *Danionella* as a novel model in the neurosciences. *Dev Dyn*. 250(4):601–611.

Doyle WI, Meeks JP. 2018. Excreted steroids in vertebrate social communication. *J Neurosci*. 38(14):3377–3387.

Dymek J, Rosenqwist G, Kuciel M, Lauriano ER, Capillo G, Zaccone G, Żuwała K. 2021. Micro- and macro-morphology of the olfactory organ of *Syngnathus typhle* (Syngnathidae, Actinopterygii). *Acta Zool*. 102(2):206–214.

Fricke R, Eschmeyer WN, Van der Laan R. 2021. Eschmeyer's Catalog of Fishes: Genera, species, references. [cited 2021 Feb 19] Available from: http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp.

Friedrich RW, Korsching SI. 1997. Combinatorial and chemotopic odorant coding in the Zebrafish olfactory bulb visualized by optical imaging. *Neuron* 18(5):737–752.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28(23):3150–3152.

Gao S, Liu S, Yao J, Li N, Yuan Z, Zhou T, Li Q, Liu Z. 2017. Genomic organization and evolution of olfactory receptors and trace amine-associated receptors in Channel Catfish, *Ictalurus punctatus*. *Biochim Biophys Acta Gen Subj*. 1861(3):644–651.

Gracheva EO, Bagriantsev SN. 2018. Communication: potassium channels define the dialect. *Curr Biol*. 28(13):R744–R746.

Gregory TR. 2020. Animal Genome Size Database. Available from: http://www.genomesize.com.

Hansen A, Anderson KT, Finger TE. 2004. Differential distribution of olfactory receptor neurons in Goldfish: structural and molecular correlates. *J Comp Neurol*. 477(4):347–359.

Hansen A, Rolen SH, Anderson K, Morita Y, Caprio J, Finger TE. 2003. Correlation between olfactory receptor cell type and function in the Channel Catfish. *J Neurosci*. 23(28):9328–9339.

Downloaded from https://academic.oup.com/mbe/article/38/9/3742/6265480 by guest on 05 November 2021

Hansen A, Zielinski BS. 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.
Hara TJ. 1975. Olfaction in fish. *Prog Neurobiol*. 5(4):271–335.
Hughes GM, Boston ESM, Finarelli JA, Murphy WJ, Higgins DG, Teeling EC. 2018. The birth and death of olfactory receptor gene families in mammalian niche adaptation. *Mol Biol Evol*. 35(6):1390–1406.
Ishikawa SA, Zhukova A, Iwasaki W, Gascuel O. 2019. A fast likelihood method to reconstruct and visualize ancestral scenarios. *Mol Biol Evol*. 36(9):2069–2085.
Jiang H, Du K, Gan X, Yang L, He S. 2019. Massive loss of Olfactory Receptors but not Trace Amine-Associated Receptors in the world's deepest-living fish (*Pseudoliparis swirei*). *Genes* 10(11):910.
Kasumyan A. 2004. The olfactory system in fish: structure, function, and role in behavior. *J Ichthyol*. 44:S180–S223.
Katoh K, Kuma K-I, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 33(2):511–518.
Khan I, Yang Z, Maldonado E, Li C, Zhang G, Gilbert MTP, Jarvis ED, O'Brien SJ, Johnson WE, Antunes A. 2015. Olfactory receptor subgenomes linked with broad ecological adaptations in Sauropsida. *Mol Biol Evol*. 32(11):2832–2843.
Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol*. 34(7):1812–1819.
Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23(1):127–128.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
Lindqvist C, Sundin J, Berglund A, Rosenqvist G. 2011. Male broad-nosed Pipefish *Syngnathus typhle* do not locate females by smell. *J Fish Biol*. 78(6):1861–1867.
Liu A, He F, Shen L, Liu R, Wang Z, Zhou J. 2019. Convergent degeneration of olfactory receptor gene repertoires in marine mammals. *BMC Genomics*. 20(1):977.
Lv L-Y, Liang X-F, He S. 2019. Genome-Wide Identification and Characterization of Olfactory Receptor Genes in Chinese Perch, *Siniperca chuatsi*. *Genes*. 10(2):178.
Michel WC, Lubomudrov LM. 1995. Specificity and sensitivity of the olfactory organ of the Zebrafish, *Danio rerio*. *J Comp Physiol A*. 177(2):191–199.
Musilova Z, Cortesi F, Matschiner M, Davies WIL, Patel JS, Stieb SM, de Busserolles F, Malmstrøm M, Tørresen OK, Brown CJ, et al. 2019. Vision using multiple distinct rod opsins in deep-sea fishes. *Science* 364(6440):588–592.
Nei M, Niimura Y, Nozawa M. 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet*. 9(12):951–963.
Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32(1):268–274.
Niimura Y. 2009. On the origin and evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 Chordate species. *Genome Biol Evol*. 1:34–44.
Niimura Y, Matsui A, Touhara K. 2018. Acceleration of olfactory receptor gene loss in primate evolution: possible link to anatomical change in sensory systems and dietary transition. *Mol Biol Evol*. 35(6):1437–1450.
Niimura Y, Matsui A, Touhara K. 2014. Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. *Genome Res*. 24(9):1485–1496.
Niimura Y, Nei M. 2005. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proc Natl Acad Sci U S A*. 102(17):6039–6044.
Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35(3):526–528.
Policarpo M, Fumey J, Lafargeas P, Naquin D, Thermes C, Naville M, Dechaud C, Volff J-N, Cabau C, Klopp C, et al. 2021. Contrasting gene decay in subterranean vertebrates: insights from cavefishes and fossorial mammals. *Mol Biol Evol*. 38(2):589–605.
Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 5(3):e9490.
Rabosky DL, Chang J, Title PO, Cowman PF, Sallan L, Friedman M, Kaschner K, Garilao C, Near TJ, Coll M, et al. 2018. An inverse latitudinal gradient in speciation rate for marine fishes. *Nature* 559(7714):392–395.
Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*. 3(2):217–223.
Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*. 16(6):276–277.
Sato Y, Miyasaka N, Yoshihara Y. 2007. Hierarchical regulation of odorant receptor gene choice and subsequent axonal projection of olfactory sensory neurons in Zebrafish. *J Neurosci*. 27(7):1606–1615.
Schulze L, Henninger J, Kadobianskyi M, Chaigne T, Faustino AI, Hakiy N, Albadri S, Schuelke M, Maler L, Del Bene F, et al. 2018. Transparent *Danionella translucida* as a genetically tractable vertebrate brain model. *Nat Methods*. 15(11):977–983.
Sharma K, Syed AS, Ferrando S, Mazan S, Korsching SI. 2019. The chemosensory receptor Repertoire of a true shark is dominated by a single olfactory receptor family. *Genome Biol Evol*. 11(2):398–405.
The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 47:D506–D515.
Tung Ho LS, Ané C. 2014. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst Biol*. 63(3):397–408.
Tyler JC. 1980. Osteology, phylogeny, and higher classification of the fishes of the order Plectognathi (Tetraodontiformes). U.S. Department of Commerce. NOAA Technical Report, NMFS Circular, vol. 434. p. 1–422.
Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*. 35(3):543–548.
Ye J, McGinnis S, Madden TL. 2006. BLAST: improvements for better sequence analysis. *Nucleic Acids Res*. 34(Web Server issue):W6–W9.
Yoshihara Y. 2014. Zebrafish olfactory system. In: Mori K, editor. The olfactory system: from odor molecules to motivational behaviors. Tokyo: Springer Japan. p. 71–96.
Zhu G, Wang L, Tang W, Wang X, Wang C. 2017. Identification of olfactory receptor genes in the Japanese Grenadier Anchovy *Coilia nasus*. *Genes Genomics*. 39(5):521–532.

Downloaded from https://academic.oup.com/mbe/article/38/9/3742/6265480 by guest on 05 November 2021