Evolution of brain-expressed biogenic amine receptors into olfactory trace amine-associated receptors

Lingna Guo\textsuperscript{1,2,4,5,\#}, Wenxuan Dai\textsuperscript{1,2,\#}, Zhengrong Xu\textsuperscript{1,2,6,7,\#}, Qiaoyi Liang\textsuperscript{8}, Eliot T. Miller\textsuperscript{9}, Shengju Li\textsuperscript{1,2}, Xia Gao\textsuperscript{6,7}, Maude W. Baldwin\textsuperscript{8}, Renjie Chai\textsuperscript{4,5,7,*}, and Qian Li\textsuperscript{1,2,3,*}

\textsuperscript{1}Center for Brain Science, Shanghai Children's Medical Center, School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, China; \textsuperscript{2}Department of Anatomy and Physiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China; \textsuperscript{3}Shanghai Research Center for Brain Science and Brain-Inspired Intelligence, Shanghai 201210, China; \textsuperscript{4}State Key Laboratory of Bioelectronics, Department of Otolaryngology Head and Neck Surgery, Zhongda Hospital, School of Life Sciences and Technology, Jiangsu Province High-Tech Key Laboratory for Bio-Medical Research, Southeast University, Nanjing 210096, China; \textsuperscript{5}Co-Innovation Center of Neuregeneration, Nantong University, Nantong 226001, China; \textsuperscript{6}Department of Otolaryngology Head and Neck Surgery, Affiliated Drum Tower Hospital of Nanjing University Medical School, Jiangsu Provincial Key Medical Discipline (Laboratory), Nanjing 210008, China; \textsuperscript{7}Research Institute of Otolaryngology, Nanjing 210008, China; \textsuperscript{8}Evolution of Sensory Systems Research Group, Max Planck Institute for Ornithology, Seewiesen, Germany; \textsuperscript{9}Macaulay Library, Cornell Lab of Ornithology, Ithaca, NY, USA

\#These authors contributed equally to this work.
*Correspondence to: renjiec@seu.edu.cn, liqian@shsmu.edu.cn

Competing interests
The authors declare no competing interests.
Abstract

The family of trace amine-associated receptors (TAARs) is distantly related to G protein-coupled biogenic aminergic receptors. TAARs are found in the brain as well as in the olfactory epithelium where they detect biogenic amines. However, the functional relationship of receptors from distinct TAAR subfamilies and in different species is still uncertain. Here, we perform a thorough phylogenetic analysis of 702 TAAR-like (TARL) and TAAR sequences from 48 species. We show that a clade of Tarl genes has greatly expanded in lampreys, while the other Tarl clade consists of only one or two orthologs in jawed vertebrates and is lost in amniotes. We also identify two small clades of Taar genes in sharks related to the remaining Taar genes in bony vertebrates, which are divided into 4 major clades. We further identify ligands for 61 orphan TARLs and TAARs from sea lamprey, shark, ray-finned fishes, and mammals, as well as novel ligands for two 5-hydroxytryptamine receptor 4 (HTR4) orthologs, a serotonin receptor subtype closely related to TAARs. Our results reveal a pattern of functional convergence and segregation: TARLs from sea lamprey and bony vertebrate olfactory TAARs underwent independent expansions to function as chemosensory receptors, while TARLs from jawed vertebrates retain ancestral response profiles and may have similar functions to TAAR1 in the brain. Overall, our data provide a comprehensive understanding of the evolution and ligand recognition profiles of TAARs and TARLs.

Keywords: trace amine-associated receptor, olfactory receptor, GPCR, receptor evolution, receptor deorphanization, site-directed mutagenesis, homology modeling
Introduction

Trace amine-associated receptors (TAARs) form a distinct subfamily of G protein-coupled receptors (GPCRs) that are specialized to detect both endogenous and exogenous biogenic amines. TAARs were initially discovered as receptors for a number of trace amines that are structurally similar to monoamine neurotransmitters (e.g., serotonin, dopamine, and histamine) but present in trace concentrations in the brain (Borowsky, et al. 2001; Bunzow, et al. 2001). The subsequent studies revealed a functional dichotomy of TAARs, with TAAR1 expressed in the brain and sensing endogenous trace amines, and all of the other TAARs highly expressed in the main olfactory epithelium (MOE) recognizing exogenous biogenic amines (Liberles and Buck 2006; Dewan 2020; Xu and Li 2020). Several TAARs are also distributed throughout the body and are found in other tissues such as kidney, heart, and testes, albeit with lower expression levels (Gainetdinov, et al. 2018). Non-olfactory TAAR1 negatively regulates excitability and monoamine neurotransmitter transmission of dopaminergic and serotonergic neurons. Psychostimulant abuse-related behaviors are reduced by TAAR1 activation and enhanced by TAAR1 knockout in mice (Bradaia, et al. 2009; Revel, et al. 2011; Achat-Mendes, et al. 2012; Liu, et al. 2020). Olfactory TAARs, on the other hand, are crucial for perception of amines, an ecologically important class of odorants. Amines are produced by amino acid decarboxylation during decomposition of proteins, and are often enriched in animal body fluids. Therefore, they have been proposed to mediate intra- and interspecific communication through the TAAR olfactory subsystem. For instance, trimethylamine is a species- and sex-specific urine odor that activates TAAR5 to attract mice and repel rats (Li, et al. 2013; Saraiva, et al. 2016). The TAAR4 ligand 2-phenylethylamine is enriched in predator urine and elicits innate avoidance behaviors in rodents (Ferrero, et al. 2011; Dewan, et al. 2013). Cadaverine can be produced from decaying animal carcasses and activates TAAR13c, triggering aversive behavior in zebrafish (Hussain, et al. 2013). Thus, TAARs play important roles in both psychostimulant addiction and social behaviors.

Since Taar genes were first cloned in 2001, they have been identified in several vertebrate genomes (Hussain, et al. 2009; Gao, et al. 2017; Eyun 2019; Sharma, et al. 2019; Dieris, et al. 2021). The numbers of intact Taar genes vary across species, ranging from 0 in dolphins, 6 in humans, 15 in mice to 112 in zebrafish. TAARs are distantly related to classic biogenic amine receptors, including dopamine and serotonin receptors. Phylogenetic studies have indicated that the 5-hydroxytryptamine receptor 4 (HTR4), a serotonin receptor subtype, is more closely related to the TAAR subfamily than other biogenic amine receptors (Hashiguchi and Nishida 2007; Hussain, et al. 2009; Li and Liberles 2016; Dieris, et al. 2021). However, there has been controversy over the exact timing of the origin of Taar genes. Initial reports suggested that the Taar gene family emerged in jawless vertebrates such as sea lamprey.
(Hashiguchi and Nishida 2007; Libants, et al. 2009). Other studies suggested a later origin in early jawed fish (Hussain, et al. 2009; Eyun, et al. 2016); because so-called sea lamprey TAARs form a monophyletic clade and lack the classic TAAR motif in the transmembrane α-helix VII, they were named Taar-like (Tarl) genes. The earliest-branching Taar genes with the intact TAAR motif are found in cartilaginous fishes, the earliest-branching extant jawed vertebrates, including elephant shark, catshark, white shark, and whale shark (Hussain, et al. 2009; Marra, et al. 2019; Sharma, et al. 2019). Recently, a comprehensive phylogenetic analysis of Tarl and Taar genes places lamprey Tarls as sister to other Tarl genes (called the Taar V subfamily in previous studies) (Hashiguchi and Nishida 2007; Eyun, et al. 2016), which together are sister to the classical Taars (Dieris, et al. 2021).

Apart from the Tarl clade, classical TAAR receptors are classified into 3 major clades (called class I, II, and III by Hussain, et al. 2009, or called clade I, II, and III by Li, et al. 2015). The tetrapod TAARs are grouped in clade I and II, while clade III is teleost-specific. Interestingly, almost all of the clade I and II TAARs possess the canonical amine-recognition motif (Asp^{3.32}; Ballesteros-Weinstein indexing for GPCRs). This motif is lost in the vast majority of clade III TAARs, which evolve the non-canonical amine-recognition motif, Asp^{5.42}. Furthermore, a few TAARs contain both Asp^{3.32} and Asp^{5.42}, and recognize diamines (which have two amino groups) (Li, et al. 2015). However, with the exception of receptors from a few model species, most of the TAARs in different species remain orphan receptors, which restricts the functional analyses of this specific family of receptors.

Here, we sought to investigate the evolutionary history and functional responses of TAARs, TARLs, and HTR4s. In total, we retrieved 702 TAAR and TARL sequences from 48 vertebrate species and constructed a phylogenetic tree. Among several aminergic receptors selected as outgroups, HTR4s are the closest relative to TAARs and TARLs. TARLs are grouped into two subfamilies. One subfamily expands greatly in lampreys, while the other one exhibits very little duplication in jawed vertebrates and appears to be lost in amniotes. We also found that TAARs can be clustered into 4 distinct clades – Ia, Ib, II, and III. To further analyze their functional relationships, we next identified novel ligands for 61 TAARs and TARLs as well as 2 HTR4s. Surprisingly, HTR4s, TARLs from jawed fishes and non-olfactory TAAR1 have similar ligand response profiles, while some lamprey TARLs recognize ligands that are also recognized by olfactory TAARs, albeit with a distinct structural basis. Consistent with the distinct functional profile of the two TARL subfamilies, we found that TARLs in jawed fishes such as zebrafish are expressed in the brain, whereas lamprey TARLs are expressed in the MOE. In sum, our comprehensive analysis of the evolution and function of TAARs/TARLs uncovers the evolutionary transitions from brain-expressed biogenic amine receptors to olfactory amine
detectors. Our study provides evidence for functional distinction and convergence between TAARs and TARLs.

**Results**

**Evolutionary history of HTR4s, TARLs, and TAARs**

To explore the evolutionary relationships of HTR4s, TARLs, and TAARs, we retrieved a thorough collection of sequences from different species. These species included 2 jawless fishes (hagfish and sea lamprey), 3 cartilaginous fishes (ghost shark, bamboo shark, and whale shark), 7 teleost fishes (medaka, catfish, goldfish, pufferfish, stickleback, fugu, and zebrafish), 1 holost fish (spotted gar), 1 lobe-finned fish (coelacanth), 1 amphibian (clawed frog), 7 reptiles, and 26 mammalian species. In total, 17 sequences of HTR4s, and 702 homologous sequences of TARLs or TAARs from 48 species were obtained (supplementary table S1, supplementary data S1). In addition, we selected 39 classical biogenic amine receptors as outgroups, consisting of 7 α adrenergic receptors, 3 β adrenergic receptors, 4 dopamine receptors, 9 muscarinic acetylcholine receptors, 7 histamine receptors, and 9 serotonin receptors from whale shark, zebrafish, coelacanth, and mouse.

The phylogenetic analyses of the receptor sequences were performed using maximal likelihood (ML) in IQ-TREE (supplementary fig. S1A-B) and nodal support was assessed both by 5,000 ultrafast bootstrap replicates (UF) (Hoang, et al. 2018) as well 100 standard bootstrap replicates; similar tree topologies and support values were obtained from multiple independent runs. Consistent with previous results, we recover monophyletic clades of HTR4s, lamprey TARLs, and TAARs with maximal support (fig. 1A, supplementary fig. S1A-B). Like in lamprey TARLs, the classical TAAR motif is also missing in TARLs from this clade (fig. 1B). To differentiate between those TARLs from the jawed vertebrates and the lamprey TARLs, we renamed the lamprey TARLs as TARLLs (the last L indicating lamprey). The TARLLs show a large expansion, consisting of 33 members in sea lamprey that can be further subdivided into 4 subfamilies, TARLL1-4 (supplementary fig. S1A). By contrast, there appeared to be generally only 1 TARL in most of jawed vertebrates (TARL1 in cartilaginous fishes, ray-finned fishes, coelacanth, and amphibians) and 2 TARLs in some species (TARL1 and TARL2 in whale shark and coelacanth, supplementary fig. S1A). We did not find any TARLs in amniotes (supplementary table S1). TARLLs and TARLs therefore display distinct diversification patterns in jawless and jawed vertebrates, suggesting distinct functional roles.

Next, we investigated the evolutionary dynamics of TAARs. TAARs form a large monophyletic clade, including receptors from species ranging from jawed fishes, amphibians, reptiles, and mammals. The TAARs can be further segregated into four major monophyletic groups or
clades (Ia, Ib, II, and III), largely corresponding to previously defined groups (Hussain, et al. 2009; Li, et al. 2015). A previous study investigated around 300 TAAR sequences and segregated them into 3 clades designated class I, II, and III (Hussain, et al. 2009). In our study, with an expanded set of species analyzed, class I TAARs were separated into two clades, which we named clade Ia and Ib. Furthermore, Hussain et al. subdivided TAARs into 28 different subfamilies (TAAR1-28). Here we found several TAARs that could not be included into the existing 28 subfamilies, so we applied new subfamily names to those TAARs (subfamilies are distinguished if they form separate groups inside the clades). For clade III teleost TAARs, we additionally named three subfamilies (TAAR29, 30, and 31). For other TAARs, we named the subfamilies by the main species in which they occur, including TAARS1-2 (S for shark), TAARC1-3 (C for coelacanth), TAARSG1-6 (SG for spotted gar), TAARR1 (R for reptile), and TAARM1-2 (M for metatherian) (fig. 1A). Interestingly, TAARs from cartilaginous fishes display a distinct phylogenetic distribution. Cartilaginous fish TAARs form two subfamilies, TAARS1 and TAARS2. TAARS1 is sister to the remaining clade Ia TAARs, whereas TAARS2 is sister to all subsequent TAARs, including clade Ib, II, and III TAARs (fig. 1A). This phylogenetic pattern suggests that the divergence of clade Ia and other TAARs (clade Ib, II, and III) occurred prior to the divergence of cartilaginous fish and bony fish.

The characteristic fingerprint motif (NSxxNPxxYxxxYxWF, where x represents non-conserved amino acids) in the transmembrane α-helix VII was identified for TAARs (Hussain, et al. 2009). We analyzed the consensus motifs in TARLLs, TARLs, and the 4 clades of TAARs (supplementary fig. S2A). In agreement with the previous study, the TAAR motif is well conserved in all of the 4 clades of TAARs. In TARLL and TARLs, we identified similar but distinct fingerprint motifs. TARLLs have a consensus motif of SxxNPxLxxxxNxxF, while TARLs have a consensus motif of NSxxNPpLYxxxxSF (fig. 1B). We also identified several conserved amino acids of TARLs and TARLLs that are absent in TAARs. Most of those are in transmembrane or extracellular regions, suggesting a role in ligand recognition, distinct from that seen in TAARs (fig. 1C).

We further analyzed the presence of pseudogenes in several species. We did not find any TAAR/TARL/TARLL pseudogenes in hagfish, confirming that TARLLs specifically emerge in lamprey. The percentages of pseudogenes differ across species, varying from 6% in mouse to 44% in whale shark (supplementary fig. S2B).

**Cell-based high-throughput screening to deorphanize receptors**

Next, we conducted high-throughput screening to further explore the functional properties of these aminergic receptors. We cloned 248 TAARs from 11 mammals (human, mouse, rat,
guinea pig, hamster, cat, horse, rabbit, sheep, rhesus, and wild boar), 2 reptiles (alligator and chicken), 6 teleosts (zebrafish, catfish, fugu, medaka, salmon, and pufferfish), 1 lobed-finned fish (coelacanth), 1 holost fish (spotted gar), and 1 cartilaginous fish (bamboo shark) (fig. 2A). We also cloned 8 TARLs from coelacanth, zebrafish, catfish, fugu, medaka, spotted gar, shark, and 28 TARLLs from sea lamprey (supplementary table S2). In addition, we cloned 2 HTR4s from mouse and zebrafish. We then challenged all of the cloned receptors with a ligand library comprised of 97 amines at 100 µM. In total, we successfully identified novel ligands for 50 TAARs, 8 TARLs, 3 TARLLs, and 2 HTR4s. The deorphaned receptors are spread across the 8 major clades (fig. 2B, supplementary table S3). For those receptors showing high responses, we subsequently performed dose-response curves. Most of the ligands activate the receptors with half maximal effective concentrations (EC$_{50}$) of 1-100 µM, and a few ligands show extremely sensitive activation with an EC$_{50}$ of 20 pM (supplementary table S4).

To better understand the ligand response profiles, we classified the 97 amines into 8 chemical clusters according to their chemical features, including predefined atom symbols, bond types, atom environment properties, and atom properties (supplementary fig. S3). The largest 3 clusters consist of a group mainly composed of primary or secondary amines (cluster 2), another group is primarily composed of amines with aromatic rings (cluster 3), and cluster 5 is mainly composed of tertiary amines. The other 5 clusters include cluster 1 (taurine), cluster 4 (creatinine), cluster 6 (isopropylamine), cluster 7 (aniline) and cluster 8 (indole and its derivatives). It is worth noting that most biogenic amine neurotransmitters are included in chemical cluster 3, and most of the other known TAAR ligands are included in cluster 2 and 5. The 52 identified ligands are mainly in cluster 2, 3, 5, 6, 7, and 8. All of the molecules are positively charged in physiological conditions, except for molecules in cluster 7 and cluster 8, which are neutral in physiological conditions.

**Functional relationships of HTR4s, TARLLs, TARLs, and TAARs**

Next, we combined our data with known ligands of TAARs from model organisms, such as human, mouse, rat, and zebrafish (Xu and Li 2020). Comparison of ligand responses showed that in general, receptors in different clades have distinct ligand profiles, responding to chemicals in distinct clusters, while receptors within the same clades tend to detect ligands from the same chemical clusters (fig. 3A). Further, TAAR and TARL orthologs from different species generally detect very similar ligands, suggesting conservation of receptor function across species (fig. 3A). However, we did observe broader ligand profiles for receptor orthologs in some species, although we could not rule out the possibility that receptor orthologs may not function equally well in vitro. It is also worth noting that TAAR-expressing
olfactory sensory neurons generally recognize the same ligands identified by in vitro assays, yet with enhanced sensitivity and increased number of ligands, likely due to optimized signaling components and chaperons in neurons in vivo (Pacifico, et al. 2012; Zhang, et al. 2013).

To better illustrate the functional relevance of HTR4s, TARLLs, TARLs, and TAARs, we performed phylogenetic principal component analysis (phyloPCA) (Revell and Collar 2009) on all of the deorphaned receptors based on the ligand profiles and the ML phylogenetic tree. In our primary dataset, since the top 6 principal components (PCs) together only explain 63.9% of the variance (supplementary fig. S4A-C), we further applied dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP) on all phyloPCA scores for better visualization (fig. 3B). To quantitatively classify receptors by their phyloPCA scores on all PC axes, we used a clustering algorithm (Partition Around Medoids, PAM) after identifying the optimal number of groups (Maechler, et al. 2019; Kassambara and Mundt 2020). We found that all of these receptors were assigned into 2 groups by PAM (supplementary fig. S4D), overlapping with the UMAP visualization (fig. 3B). We noted that almost all of the receptors from clade Ia TAARs (except zebrafish TAAR10b and medaka TAAR21e), TARLs, and HTR4s are clustered in PAM group 2, suggesting that they share similar ligand response profiles. Some clade Ib TAARs (TAAR4 and TAAR12) and a clade III TAAR (catfish TAAR13g) also appear in the UMAP plot close to PAM group 2, which might be due to their responses to some ligands in cluster 3 (amines with aromatic rings). However, most clade Ib, II, and III TAARs cluster in PAM group 1 together with shark TAARS1b both in the PAM clustering and the UMAP plot. For the 4 deorphaned sea lamprey TARLLs, 3 (TARLL2a, TARLL3h, and TARLL4a) are clustered into PAM group 1, and only TARLL1b is clustered into PAM group 2. Since Dieris, et al. 2021 showed a different topology on the clades of TARLL, TARL, and TAAR with higher support values, we also examined the effect of the difference in topologies by analyzing responses of a reduced set of receptors present in both phylogenies. The clustering results from the topology comparison are similar, and are consistent with the results from the full dataset (supplementary fig. S5A-D).

TAARs show distinct ligand binding features
We identified novel ligands for 1 TAARS1 (TAARS1b), 5 clade Ia, 10 clade Ib, 24 clade II, and 10 clade III TAARs from 18 different species (supplementary table S3), providing an opportunity to understand TAAR-ligand interaction in depth. We previously reported that clade III TAARs evolved the non-canonical Asp$^{5.42}$ to detect amine ligands, while other TAARs utilize the conserved Asp$^{3.32}$ to detect amine ligands (Li, et al. 2015). We observed the same pattern in the current study (supplementary fig. S6A). Further, we found that Asp$^{3.32}$ (but not Asp$^{5.42}$)
is well conserved in sequences of HTR4, TARLL, and TARL, implying that HTR4, TARLL, and TARL use the conserved Asp$^{3.32}$ to detect amines. Interestingly, a few clade II mammalian TAARs, including TAAR6 and TAAR8, have Asp$^{5.43}$ instead of Asp$^{5.42}$. This results from the loss of an amino acid at 5.44 in the fifth transmembrane segments of those receptors (supplementary fig. S6A). In spite of this fact, simulations do suggest that those TAARs may recognize diamines according to computational simulations (Izquierdo, et al. 2018). However, our results clearly showed that Asp$^{5.43}$-containing TAAR6 and TAAR8 family members, including TAAR6 from mouse, rat, cat, TAAR8a from rat, wild boar, and TAAR8b from rabbit, can only be activated by monoamines other than diamines (supplementary fig. S6B, supplementary table S3). We suspect that Asp$^{5.43}$ might be directed away from the potential binding pocket, and thus may not be involved in amine recognition.

Our previous study also proposed a two-step model for Asp$^{5.42}$ acquisition, suggesting that an ancestral TAAR of clade III TAARs acquired Asp$^{5.42}$, gaining diamine sensitivity, and subsequently lost Asp$^{3.32}$ (Li, et al. 2015). Interestingly, we found that both Asp$^{3.32}$ and Asp$^{5.42}$ are present in the cartilaginous fish TAARS2 subfamily (supplementary fig. S6A). The facts that TAARS2 already has both Asp$^{3.32}$ and Asp$^{5.42}$ together with the observation that some clade II mammalian TAARs contain both Asp$^{3.32}$ and Asp$^{5.43}$ suggest two alternative scenarios. In one scenario, these two residues may have arisen independently and convergently multiple times, in shark TAARS2, and in some clade II and clade III TAARs. Alternatively, after the divergence of clade Ia, the ancestor of the subsequent TAARs may have evolved both Asp$^{3.32}$ and Asp$^{5.42}$. Asp$^{5.42}$ may have then been lost in clade Ib TAARs but retained in the common ancestor of clade II and III TAARs, followed by subsequent loss of Asp$^{5.42}$ in some clade II TAARs or shifting to Asp$^{5.43}$ in other clade II TAARs (as a result of a deletion, supplementary fig. S6A), and loss of Asp$^{3.32}$ in the majority of clade III TAARs (supplementary fig. S6C).

In addition, it was previously reported that clade Ib TAAR2-4 mainly recognize primary and secondary amines, but clade II TAAR5-9 mainly recognize tertiary amines (Ferrero, et al. 2012). Here we found that clade II TAARs can also recognize a relatively large number of cluster 2 ligands with primary and secondary amines. This is consistent with previous studies showing that TAAR3 and TAAR4 are more broadly tuned in vivo, and could recognize tertiary amines as well (Zhang, et al. 2013; Dewan, et al. 2018). We also found that clade Ib TAARs respond predominantly to cluster 3 ligands (amines with aromatic rings), and clade II TAARs respond predominantly to cluster 5 ligands (tertiary amines) (supplementary fig. S7A-B).

Another interesting finding is the functional convergence between shark TAARS1 and TAAR5. Shark TAARS1 appears closely related to clade Ia TAARs according to the phylogenetic tree.
However, the ligand profile of TAARS1 is more similar to TAAR5 in clade II than to TAARs in clade Ia, which mainly recognizes cluster 3 molecules (amines with aromatic rings). Bamboo shark TAARS1b can recognize 1 ligand from cluster 2 (primary or secondary amines), 1 ligand from cluster 3 (amines with aromatic rings), and 5 ligands from cluster 5 (tertiary amines), which are the ligands of TAAR5 as well (fig. 3A-B, supplementary fig. S8A). We used the R package l1ou (Khabbazian, et al. 2016) to search across the phylogenetic tree for well-supported shifts to new optima in ligand response profiles (supplementary fig. S8B). A notable shift was indeed discovered in shark TAARS1b following the divergence from clade Ia TAARS (supplementary fig. S8B). This functional convergence of TAARS1 and TAAR5 is intriguing: as TAAR5 ligands are potent allomones in mammals (Li, et al. 2013), the ecological relevance of TAARS1b ligands for cartilaginous fishes will be interesting to explore.

Independent development and functional convergence of TARLL and TAAR

Next, we examined the TARLL clade in more detail. Previous studies have suggested that TARLLs are expressed in the MOE and mediate olfactory function in the sea lamprey (Scott, et al. 2019). However, detailed analyses of those receptors are still lacking. Here, we successfully deorphaned 3 lamprey TARLLs (TARLL1b, 3h, 4a) out of 28 cloned and functionally tested receptors, and further investigated the structural basis in TARLLs for ligand recognition (supplementary table S3). Expression patterns of the 3 related Tarll genes determined by in situ hybridization in the MOE of sea lamprey also supports the hypothesis that lamprey TARLLs mainly function as olfactory receptors (supplementary fig. S9A). TARLL4a has a broad ligand profile of amines with 8 ligands from cluster 2 (primary or secondary amines) and 1 ligand from cluster 6 (isopropylamine). The responses of TARLLs to amines are very specific as they are not responsive to other chemicals including aldehydes, ketones, acids, esters, and alcohols (fig. 4A). The ligand profile includes long alkyl chain amines, such as hexylamine and heptylamine, and amines whose nitrogen is connected to a secondary carbon, such as isopropylamine and cyclohexylamine (fig. 4B, supplementary fig. S9B). This suggests that the binding pocket of TARLL4a is flexible in its volume to fit the sizes of a variety of ligands.

Interestingly, TARLL3h recognizes tetramethyl-1,4-butanediamine and N,N,N’,N’-tetramethyl-1,3-propanediamine, diamines with two amino groups (fig. 4C). Recognition of diamines in TAARs involves two aspartic acids, Asp³.32 and Asp⁵.42 (Li, et al. 2015; Sharma, et al. 2016). However, although Asp³.32 is retained in TARLL3h, Asp⁵.42 is not found, implying a distinct diamine recognition mechanism in TARLL3h. To investigate this further, we performed homology modelling and ligand docking in TARLL3h. In addition to Asp³.32 that forms a salt bridge with N,N,N’,N’-tetramethyl-1,4-butanediamine or N,N,N’,N’-tetramethyl-1,3-
propanediamine, Tyr$^{3.33}$ binds to the other amino group through pi-cation interactions in TARLL3h (fig. 4D). Notably, the Tyr$^{3.33}$ site is one of the well-conserved amino acids in the TARLL clade (fig. 1C). To verify the docking results, we performed site-directed mutagenesis in TARLL3h. Mutation of Asp$^{3.32}$ or Tyr$^{3.33}$ to alanine (D3.32A or Y3.33A) completely eliminated the receptor activity, while mutation of Tyr$^{3.33}$ to phenylalanine (Y3.33F) retained pi-cation interactions and the receptor activity yet resulted in slightly reduced affinity ($EC_{50}$ three times higher than wild type receptor) (supplementary fig. S9C). These results suggest the potential roles of Asp$^{3.32}$ and Tyr$^{3.33}$ of TARLL3h in diamine recognition.

Moreover, TARLLs seem to extend ligand profiles beyond the profiles seen in TAARs. TARLL1b can be activated by indole and its derivatives, which are not ligands for any TAARs (fig. 4E, supplementary fig. S9D-E). Indole and its derivatives are neutral in physiological conditions, and are found to activate members of odorant receptor (OR) family (Pfister, et al. 2020). To probe the structural basis of indole recognition, we performed homology modelling and ligand docking in TARLL1b (fig. 4F). The docking results suggested that Phe$^{5.38}$ is the critical site to stabilize indole through pi-pi interactions with the aromatic rings of indole. This site is also important for recognition of indole derivatives such as 1-methylindole (fig. 4F).

**TARL and TAAR1 preserve the ligand profiles of HTR4 and independently gain new functions**

As underscored by the UMAP plot (fig. 3b), TARL, HTR4, and TAAR1 appear to have similar ligand profiles and are cluster together (PAM group 2). Most of their ligands belong to cluster 3 ligands (amines with aromatic rings), which are mainly neurotransmitters and their derivatives (fig. 3A). In addition, the analysis of expression levels in a panel of zebrafish tissues demonstrates that these receptors are mainly expressed in the brain, instead of in the MOE where other TAARs and TARLLs are expressed (Liberles and Buck 2006; Scott, et al. 2019) (fig. 5A). The expression pattern and ligand profile of TARLs suggest a possible function as a neurotransmitter receptor in the brain rather than as peripheral chemosensors.

To better compare their ligand profiles, we selected the 16 ligands that activate HTR4, TARL, and TAAR1 in our screening assay (supplementary table S3). This set included 3 molecules from chemical cluster 2 (primary or secondary amines), 11 molecules from chemical cluster 3 (amines with aromatic rings), and 2 molecules from chemical cluster 5 (tertiary amines). We then compared the responses of HTR4 from mouse and zebrafish, TARL1 from zebrafish, catfish, fugu, medaka, spotted gar, and coelacanth, TARL2 from coelacanth and shark, TAAR1 from mouse and zebrafish to the 16 ligands at 10 µM (fig. 5B). In general, all receptors showed high activation levels to the 4 HTR4 ligands, suggesting functional conservation with
HTR4. Furthermore, they also displayed distinct ligand profiles. TARL1s seem to have broader ligand profiles compared to HTR4. The TARL1 ligands include not only tryptamine and its derivatives, but also phenylethylamine and its derivatives, which are also brain neurotransmitters but have a totally different aromatic ring structure from tryptamine. Therefore, TARL1 appears not only to preserve the possible ancestral function shared with HTR4, but also gains new functions. In contrast, TARL2s cannot recognize phenylethylamine and its derivatives and have a similar ligand profile to HTR4. Consistent with these results, the phyloPCA analyses showed that TARL1s are clearly separated from HTR4s and TARL2s (fig. 5C), and a functional shift was also observed in TARL1s after divergence from TARL2s (supplementary fig. S8B, S10A-B). TAAR1s share a similar ligand profile to HTR4 and TARL as a whole, but also differ across species. Mouse TAAR1 not only recognizes the ligands of HTR4 and TARL, but also has slight responses to long-chain primary monoamines from chemical cluster 2 (hexylamine, heptylamine, and octylamine). Zebrafish TAAR1 and coelacanth TAARC1b, the counterpart of TAAR1 in coelacanth, however, have the same ligand profile as HTR4 (fig. 5B). In addition, we tested dose-dependent responses of these receptors to tryptamine and serotonin, and found that the sensitivities vary for different receptors to the same ligand (supplementary fig. S11A-B).

In summary, we trace the evolution and functional differentiation of the TAAR/TARL family. HTR4 is present in jawless fishes and is retained in all vertebrates. TARL1 exists in cartilaginous fishes, teleost, holost, coelacanth, and in amphibians, but is lost in amniotes. TARL2 emerges in cartilaginous fishes, and is lost in ray-finned fishes and tetrapods (but is present in coelacanth). On the other hand, TAAR1 is found in some ray-finned fish species and is preserved in tetrapod (fig. 5D). This phylogenetic distribution, together with the common expression in the brain and the functional conservation suggest that TARL and TAAR1 may have gradually developed new functions in the brain after diverging from a common ancestor shared with HTR4.

Discussion

In this study, we systematically analyzed the evolutionary history and functional response properties of TAARs. By constructing a phylogeny of 702 TAAR and TARL sequences from 48 vertebrate species, we demonstrated that TAARs and TARLs are very likely derived from an ancestral Htr4 gene duplication. Furthermore, we found that TARLs form two monophyletic clades: one is lamprey-specific (we named TARLL) and the other includes TARLs from jawed vertebrates. For TAARs, our study classified them into 4 distinct clades – Ia, Ib, II, and III. Further ligand screening data showed that HTR4s, TARLs from jawed fishes, and the non-olfactory TAAR1 are expressed in the brain and have similar broad ligand profiles. In contrast,
lamprey TARLLs recognize distinct ligands that are also recognized by olfactory TAARs, consistent with their specific expression in the MOE. Taken together, our study outlines the evolutionary history and functional evolution of brain-expressed biogenic amine receptors into olfactory amine-detecting receptors, and suggests widespread functional diversification and convergence in TAARs and TARLs.

**Evolution of brain-expressed biogenic amine receptors into olfactory amine detectors**

Investigating the gain and loss of genes together with shifts in ligand profiles provides important insights into the origin and function of TAARs. Previous evolutionary studies were performed mainly on limited number of TAARs and TARLLs with very few functional analyses (Libants, et al. 2009; Hussain, et al. 2013; Li, et al. 2015; Scott, et al. 2019; Dieris, et al. 2021).

In our study, we performed a thorough phylogenetic analysis of 702 TAAR and TARL sequences from 48 species including jawless fishes, cartilaginous fishes, teleost fishes, holost fishes, lobe-finned fishes, amphibians, reptiles, and mammals. We further cloned and tested 286 receptors, and identified novel ligands for 61 receptors, and integrated the ligand profiles with phylogenetic analyses. Hence, our data provide the most comprehensive evolutionary and functional understanding of TAARs and TARLs to date. Furthermore, we performed sequence analysis and receptor deorphanization in several species including bamboo shark, spotted gar, and coelacanth that have not been studied before. These species are situated at important phylogenetic junctures and provide insight into the functional evolution of this clade of receptors.

Based on the results of phylogenetic analyses and functional assays, we classified the deorphaned receptors into 2 functional types (fig. 3B). One functional group is mainly composed of HTR4s, TARLs, and TAAR1. This group of receptors mainly recognizes amines from chemical cluster 3 (amines with aromatic rings) that are mostly neurotransmitters and their derivatives. The other group of receptors is composed of TARLLs, non-TAAR1 clade Ia TAARs, clade Ib, II, and III TAARs. The ligand profile of this group of receptors is broad, including cluster 2 (primary or secondary amines) and cluster 5 (tertiary amines) ligands that are mostly alkyl amines and are known olfactory odorants. Moreover, the above observations agree with the receptor expression patterns. HTR4s, TARLs, and TAAR1 are primarily expressed in the brain, while TARLLs and other TAARs are specifically expressed in the MOE. Therefore, we hypothesize that TARLLs independently expanded in sea lamprey and evolved into olfactory receptors detecting water-soluble amines. By contrast, TARLs from jawed vertebrates are strongly conserved in number and their function was eventually replaced by TAAR1 in amniotes. TAAR1s are non-olfactory, brain-expressed and play an important role in regulating psychostimulant abuse-related behaviors (Bradaia, et al. 2009; Revel, et al. 2011;
Achat-Mendes, et al. 2012; Liu, et al. 2020), and are proposed to regulate synaptic strength in different brain regions. It is possible that TARLs may function similarly in the brain of jawed vertebrates. As a side note, the distinct expression pattern of TAAR1 in the brain rather than in the MOE could be due to the insulated TAAR1 genomic domain from two newly identified olfactory TAAR enhancers (Fei, et al. 2021; Shah, et al. 2021).

Given our results, we propose that these receptors may be functionally assigned to two distinct classes: brain-expressed biogenic amine receptors and MOE-expressed olfactory amine detectors (fig. 5E). The brain-expressed biogenic amine receptors include HTR4, TARL, and non-olfactory TAAR1, and the MOE-expressed olfactory amine detectors include TARLL and olfactory TAARs. The brain-expressed biogenic amine receptors mainly recognize ligands with aromatic rings (chemical cluster 3) that are known as amine neurotransmitters or trace amines, while the MOE-expressed olfactory amine detectors mainly recognize ligands that are primary and secondary alkyl amines (chemical cluster 2) as well as tertiary amines (chemical cluster 5). Our phylogenetic analysis suggests that the MOE-expressed TAARs likely evolved from the brain-expressed biogenic amine receptors. Similar evolutionary trajectories are observed in other olfactory receptor families, such as formyl peptide receptors (FPR) and membrane-spanning 4A (MS4A), which initially function as receptors in the immune system (Greer, et al. 2016; Dietschi, et al. 2017).

Evolution of amine recognition motifs

Prior to this study, the known amine recognition sites in TAARs include Asp$^{3.32}$ which is conserved in the majority of aminergic receptors, as well the non-canonical Asp$^{5.42}$. Binding of monoamines requires either one of the two sites, while binding of diamines needs both sites (Li, et al. 2015). With more receptors deorphaned in our study, we found several exceptions to this model and described receptors with distinct amine recognition motifs, further expanding our understanding of ligand-binding mechanisms in this family.

Firstly, we show that Asp$^{3.32}$ is maintained in all TAARs and TARLs, suggesting the evolutionarily conserved monoamine recognition motif in amine-detecting receptors. Interestingly, the lamprey TARLL2 subfamily members also contain Asp$^{5.42}$ (supplementary fig. S6A). One member of TARLL2, TARLL2a (also called TAAR346a) has been deorphaned and shown to recognize cadaverine and spermine (Scott, et al. 2019). It is very likely that both Asp$^{3.32}$ and Asp$^{5.42}$ in TARLL2a contribute to diamine or polyamine recognition, in a similar way to TAARs.
Secondly, exceptions to the use of canonical amine recognition motifs exist in TARLLs. Our homology modeling data suggest that binding of indole in TARLL1b requires Phe\textsuperscript{5.38} instead of Asp\textsuperscript{3.32} (fig. 4F). The existence of other binding sites besides Asp\textsuperscript{3.32} or Asp\textsuperscript{5.42} provides a possible structural basis for a few TAARs lacking either Asp\textsuperscript{3.32} or Asp\textsuperscript{5.42} (supplementary fig. S6C). Surprisingly, we also found that TARLL3h without Asp\textsuperscript{5.42} is able to recognize diamines. TARLL3h possibly detects diamines through Asp\textsuperscript{3.32} and Tyr\textsuperscript{3.33}, representing a novel diamine-binding mechanism.

Thirdly, we identified some interesting features of amine recognition motifs in TAARs. In clade II TAARs, one amino acid at 5.44 site is lost during evolution (supplementary fig. S6A). Therefore, TAAR6 and TAAR8 subfamily members have Asp\textsuperscript{5.43} instead of Asp\textsuperscript{5.42}. Previous studies have conjectured that these mammalian TAAR6 and TAAR8 can recognize diamines by computational analysis (Izquierdo, et al. 2018). Here, we found that several TAAR6 and TAAR8 subfamily members can only detect monoamines but not diamines. Those results suggest that Asp\textsuperscript{5.43} does not function similarly to Asp\textsuperscript{5.42}, possibly due to orientation of Asp\textsuperscript{5.43} away from the binding pocket. In addition, the TAAR9 subfamily members have Asp\textsuperscript{3.32} but not Asp\textsuperscript{5.42}, yet mouse TAAR9 can recognize polyamines, including spermine, spermidine, and diamine such as cadaverine (Saraiva, et al. 2016). Our experiments also detected activation of TAAR9 from rat, cat, rabbit, and hamster by spermidine, spermine, and cadaverine (supplementary table S3), suggesting another diamine/polyamine recognition mechanism without Asp\textsuperscript{5.42} (Jia, et al. 2021). Together, the evolutionary conservation and diversification of amine recognition motifs highlights the structural versatility of those amine-detecting receptors.

**Materials and Methods**

**Phylogenetic tree construction**

The receptor sequences of human, mouse, rat, and zebrafish were retrieved in previous publication (Hussain, et al. 2009; Saraiva, et al. 2015). Among the identified species, zebrafish has the largest number of Taar genes. However, the exact number of zebrafish Taar genes is still controversial. Hussain *et al.* reported 112 Taar genes in zebrafish (Hussain, et al. 2009), whereas RNA-seq analysis of the zebrafish MOE samples revealed 118 Taar genes in another study (Saraiva, et al. 2015). To acquire a more accurate collection of the zebrafish Taar gene repertoire, we validated the sequences from the two studies by comparing them to sequences in the Ensembl database. After deleting the pseudogenes and duplicated sequences, we eventually acquired 108 functional Taar genes (supplementary table S5). The receptor sequences of other species were obtained using tblastn against genome assemblies in NCBI. Sequences with abnormal characters that indicate inaccurate sequences were deleted unless we confirmed the correct sequences by cloning. The criteria used to determine a functional
receptor include: (i) the coding sequence length longer than 750 bp; (ii) the presence of 7 transmembrane domains predicted by TMpred web server (https://embnet.vital-it.ch/software/TMPRED_form.html); (iii) position within the TARLL, TARL, or TAAR clade in the constructed phylogenetic tree (see below). The sequences of functional receptors were included in the supplementary data S1. For TAAR pseudogene identification, the numbers of pseudogenes of human, mouse, rat, horse, bat, elephant, armadillo, platypus, chicken, and zebrafish were obtained from a previous publication (Eyun, et al. 2016). The pseudogenes of alligator, coelacanth, spotted gar, whale shark, and sea lamprey were retrieved by using the protein sequences of the functional TAARs/TARLs as a query in tblastn against the genome of the corresponding species. The previously identified functional receptors and the new receptors with lengths less than 250 amino acids were removed. For the remaining receptors, additional blastp searches were performed against human, mouse, rat, and zebrafish protein databases, and the receptors were retained as potential TAAR/TARL pseudogenes only if the most homologous aligned proteins were TAARs. Pseudogenes were defined as receptors containing interrupting stop codons, frameshifts, or deletions in conserved regions. The identified pseudogenes were listed in supplementary data S2.

Multiple sequence alignment was carried out by MAFFT v7.313 using E-INS-I option and an opening gap penalty of 1.8, intending to reduce gaps as many as possible (Katoh and Standley 2013). TrimAl was applied to trim columns with gaps at a percentage over 90% (Capella-Gutierrez, et al. 2009). The phylogenetic tree was constructed by maximum likelihood (ML). The amino acid substitution model (JTT+G4) was automatically selected by ModelFinder (Kalyaanamoorthy, et al. 2017) in Phylosuite (Zhang, et al. 2020). Maximum likelihood trees were built using IQ-TREE v1.6.8 (Nguyen, et al. 2015) integrated in Phylosuite; multiple runs produced similar results. Nodal support was assessed both by 5,000 replicates of the ultrafast bootstrap approximation (UF) (Hoang, et al. 2018) implemented in IQ-TREE, as well as 100 standard bootstrap replicates. Figtree v1.4.4 was used to visualize and modify the phylogenetic tree (http://tree.bio.ed.ac.uk/software/figtree/). The species tree was generated using Timetree (http://www.timetree.org) and visualized in Figtree v1.4.4. Sequence logo figures were generated by WebLogo (Crooks, et al. 2004). Transmembrane regions were determined according to human TAAR9 predicted by GPCRdb (Pandy-Szeker, et al. 2018).

**Chemical clustering analysis**

The structures of 97 small molecules were retrieved from PubChem (https://pubchem.ncbi.nlm.nih.gov). Hierarchical clustering of these ligands was generated in the Canvas module integrated in Schrödinger Suite platform for chemical analysis (Sastry, et al. 2010). We appointed MACCS (Molecular ACCess System) binary fingerprints for each
molecule and ran the hierarchical clustering (Durant, et al. 2002). The results were displayed using the Kelley criterion (Kelley, et al. 1996) to calculate distances and generate clusters.

Homology modeling and ligand docking
Homology models of sea lamprey TARLL1b and TARLL3h were generated using GPCR-I-TASSER (Zhang, et al. 2015). The predicted models were refined to the prepared states to adopt a physiological state by Protein Preparation Wizard module integrated in Schrödinger Suite (Sastry, et al. 2013). The ligands were prepared by LigPrep of Maestro (Maestro 2020), with the states consistent with physiological pH. We then performed receptor-ligand docking in the Induced-Fit Docking module of the Schrödinger platform by generating several poses of ligand-receptor interactions (Farid, et al. 2006). The final poses were selected according to the docking score and glide model.

Chemicals
Amine compounds were purchased from Sigma-Aldrich. Non-amine compounds were kindly donated by Hanyi Zhuang’s lab (Shanghai Jiao Tong University School of Medicine). All chemicals were dissolved in distilled water, DMSO or anhydrous ethanol at the concentration of 10 mM and stored at -20 °C. The detailed information of all chemicals was included in supplementary table S6.

Cell lines
HEK293T cells used for the SEAP assay were cultured in Dulbecco’s Modified Eagle Medium (DMEM, Thermo Fisher Scientific) supplemented with 10% bovine calf serum (BCS, HyClone) and 5% penicillin-streptomycin solution (Thermo Fisher Scientific). The HEK293T-derived Hana3A cell line used for Dual-Glo luciferase assay was grown in DMEM with 10% bovine calf serum, 5% penicillin-streptomycin solution with or without 1 μg/mL puromycin (ApexBio). Both cell lines were cultured at 37 °C with 5% CO₂.

Genomic DNA used for receptor cloning
All receptor genes were cloned from genomic DNA except zebrafish Htr4 that was cloned from zebrafish brain cDNA (AB strain). Sea lamprey genomic DNA was provided by Dr. Jeramiah Smith from University of Kentucky. Fugu genomic DNA was provided by Dr. Byrappa Venkatesh from National University of Singapore. Pufferfish genomic DNA was provided by Dr. Ferenc Mueller from University of Birmingham. Medaka genomic DNA was provided by Dr. Ivan Conte from the Telethon Institute of Genetics and Medicine (TIGEM). Coelacanth genomic DNA was provided by Dr. Jeremy Johnson from the Broad Institute. Brownbanded bamboo shark and spotted gar were purchased from commercial suppliers, sacrificed
following protocols approved by the Harvard University Animal Care and Use committee (IACUC); tissue samples are accessioned in the Museum of Comparative Zoology (Ichthyology: bamboo shark #171795, spotted gar #171800). Genomic DNA of catfish, salmon, alligator, chicken, dog, rabbit, cow, guinea pig, hamster, cat, horse, sheep, wild boar, and rhesus were purchased from Zyagen. The target fragments were inserted into pcDNA3.1-(Invitrogen) or modified pcDNA3.1- that contains an addition of DNA fragments encoding the N-terminal 20 amino acids of bovine rhodopsin in the N-terminal of target fragments.

**SEAP assay**

HEK293 cells were plated onto poly-D-lysine (Sigma-Aldrich) pre-incubated 96-well plates (Corning) with 50 µL growth medium (DMEM medium with 10% BCS and 5% penicillin-streptomycin solution) and cultured for 18-24 hours. Receptor plasmids were co-transfected with Cre-SEAP plasmid using polyethyleneimine (PEI, Polysciences) and incubated at 37 °C with 5% CO₂. 4 hours later, cells were incubated with or without test compounds diluted in serum-free DMEM for 48 hours, followed by incubating for 2 hours at 70 °C. After returning to room temperature, 50 µL supernatant from each well was incubated with an equal volume of 0.3 mM 4-methylumbelliferyl phosphate (4-MUP, Sigma-Aldrich) in 2 M diethanolamine bicarbonate (Sigma-Aldrich), pH 10.0 for 15 minutes. Fluorescence was measured with a BioTek Microplate reader.

**Dual-Glo luciferase assay**

Hana3A cells were plated onto 96-well plates (Greiner bio-one) with 50 µL growth medium (DMEM medium with 10% BCS and 5% penicillin-streptomycin solution with puromycin) and cultured at 37 °C with 5% CO₂. After 18-24 hours, cells in each well were co-transfected with 50 ng receptor plasmid, 10 ng CRE-Luc, 10 ng pRL-SV40, and 10 ng mRTPs using Lipofectamine 2000 (Invitrogen), and incubated for 18 hours. Medium was then replaced by 25 µL CD293 (Invitrogen) containing 1% glutamine with or without test compounds for 4 hours. The chemiluminescence of firefly luciferase and renilla luciferase were measured with a Biotek Microplate reader using the Dual-Glo Luciferase Assay System (Promega).

**Analyses of functional assays**

We conducted phylogenetic principle component analyses (phyloPCA) on the receptor response data (either binary responses of all deorphaned receptors or continuous values of HTR4, TARL, TAARC1 and TAAR1 to 16 ligands) together with the trimmed ML trees defined above using phyl.pca() in the phytools package of R (Revell 2009). For the phyloPCA output of all receptor responses, we applied UMAP (Uniform Manifold Approximation and Projection)
on the scores of all 52 PCs to further reduce dimensions and visualize on two UMAP axes. We also used `fviz_nbclust()` in R package factoextra (Kassambara and Mundt 2020) to determine the optimal number of clusters of deorphaned receptors using all phyloPCA scores. When setting the maximum number of clusters to be less than 12, the optimal number of clusters was two. We then used PAM (Partitioning Around Medoids) to identify which receptors belonged to each of the two clusters in the R package cluster (Maechler, et al. 2019). To detect past shifts in functional profiles across the phylogeny, we applied l1ou to scores on the first 5 PCs (which explained 58.2% of the total variance) from the all-receptor phyloPCA output using the pBIC criterion (Khabbazian, et al. 2016). The l1ou approach can efficiently use multiple traits to compute the number of shifts across the phylogeny under an Ornstein-Uhlenbeck process using a LASSO (Least Absolute Shrinkage and Selection Operator) procedure (Tibshirani 1996) without any prior hypotheses regarding locations of shifts. To compare the effect of the slightly different topology reported in Dieris et al. 2021, we trimmed both topologies to include the shared taxa (34 receptors with identified ligands) and performed phyloPCA and clustering analyses as described above. l1OU analyses were performed using the scores of all PCs on both reduced datasets.

**In situ hybridization**

Sea lamprey *Tarl* anti-sense probes were designed against the entire coding region of each receptor gene. The primers used for making probes were included in supplementary table S7. Coding region fragments were amplified from TARLL plasmids, and used for synthesis of digoxigenin-labeled RNA probes with DIG RNA labeling kit (SP6/T7). Sea lamprey MOE sections were cut at 14 µm, and hybridized with RNA probes with 1:200 dilution in the hybridization solution at 58 °C overnight. On the second day, sections were washed in 5X SSC (Invitrogen) 2 times for 5 minutes and 0.2X SSC 2 times for 30 minutes at 70 °C. The sections were incubated with blocking buffer at room temperature for 1 hour, and then incubated with alkaline phosphatase-conjugated peroxidase-anti-DIG (Roche) with 1:500 dilution in the blocking buffer for 1 hour at room temperature. Finally, sections were incubated in nitro blue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate (NBT/BCIP, Thermo Fisher Scientific) with 1:50 dilution in the alkaline phosphatase buffer at room temperature until strong signals were observed. The procedures of sea lamprey handling were approved by the Michigan State University Institutional Animal Use and Care Committee (03/14-054-00 and 02/17-031-00).

**qPCR**

Zebrafish RNA of five tissues (brain, MOE, heart, kidney, and liver) were extracted from zebrafish (AB strain) with Trizol (Invitrogen), and reverse transcribed into cDNA with RT SuperMix (ApexBio). qPCR was performed using 10 µl reaction system containing SYBR
green indicator (Roche), cDNA from each tissue, qPCR primers for receptor genes. The qPCR
primers were included in supplementary table S7. Handling procedures of zebrafish were
approved by the Institute of Neuroscience, Chinese Academy of Sciences.

Acknowledgements
We thank Liang Jia and Dr. Weiming Li at Michigan State University for providing sea lamprey
MOE samples, Dr. Jiulin Du at Institute of Neuroscience, Chinese Academy of Sciences for
providing zebrafish tissues. We also would like to thank Wei Han and Dr. Suwen Zhao at
iHuman Institute, ShanghaiTech University for help in receptor homology modeling and ligand
docking. Computational analyses were performed on the O2 High-performance Computer
Cluster at Harvard Medical School supported by Dr. Stephen D. Liberles. We thank Core
Facility of Basic Medical Sciences, Shanghai Jiao Tong University School of Medicine for
imaging assistance. This work was supported by National Natural Science Foundation of
China (31771154, 32122038, and 31970933 to Q.L., 82030029 and 81970882 to R.C.), the
National Key Research and Development Program of China (2021ZD0203100), the Basic
Research Project (21JC1404500 to Q.L.) and Shanghai Brain-Intelligence Project
(18JC1420302) from the Science and Technology Commission of Shanghai Municipality,
Shuguang Program supported by Shanghai Education Development Foundation and
Shanghai Municipal Education Commission (21SG16 to Q.L.), Program for Young Scholars of
Special Appointment at Shanghai Institutions of Higher Learning (QD2018017 to Q.L.), and
Fundamental Research Funds for the Central Universities (Shanghai Jiao Tong University,
17X100040037 to Q.L.).

Author contributions
L.G., W.D., Z.X., X.G., R.C., and Q.L. designed experiments and analyzed data; W.D., QY.L.,
M.W.B., and Q.L. wrote the manuscript; L.G., W.D., Z.X., S.L., and QY.L. performed molecular
cloning, site-directed mutagenesis, SEAP assay, and luciferase assay; W.D. and Z.X.
generated homology modeling of TAARs; W.D., QY.L., Z.X., E.T.M., and M.W.B. performed
phylogenetic and comparative analyses.

Figure legends
Figure 1. Phylogenetic analysis showing the evolutionary history of HTR4, TAARs and
TAAR-like receptors (TARL and TARLL). (A) Radial layout of maximal likelihood
phylogenetic tree of HTR4, TAARs, TARLs and TARLLs. 758 amino acid sequences from 48
species were included for phylogenetic inference. Outgroups include 7 α adrenergic receptors,
3 β adrenergic receptors, 4 dopamine receptors, 9 muscarinic acetylcholine receptors, 7
histamine receptors, and 9 serotonin receptors from whale shark, zebrafish, coelacanth, and mouse. Different groups of vertebrate species are color-coded (scale bar = 0.4 substitutions per site). According to our updated nomenclature, clade Ia TAARs are mainly composed of TAARC1, TAARSG1, TAAR1 from jawed vertebrates, and teleost-specific TAAR10, 11, 21, 27 subfamilies. Clade Ib TAARs contain TAARC4, tetrapod-specific TAAR2-4, and teleost-specific TAAR12 subfamilies. Clade II TAARs are composed of TAARC2, TAARR1, TAARM1-2, and mammal-specific TAAR5-9 subfamilies. Clade III TAARs contain TAARC3, TAARSG2-6, and teleost-specific TAAR13-20, 22-26, 28-31 subfamilies. (B) Sequence logo plots of transmembrane segment VII of the 4 major TAAR clades (clade Ia, Ib, II, and III), as well as of TARLL, TARL, HTR4, and outgroup receptors. The brown amino acids represent the TAAR fingerprint motif. The green and orange amino acids represent the motifs for TARLL and TARL, respectively. The most conserved proline at 7.50 site in the Ballesteros-Weinstein indexing system is also labeled. (C) Conserved sites in TARLL and TARL subfamilies shown in the transmembrane segments and extracellular loops are shown.

**Figure 2. Summary of deorphaned receptors.** (A) The evolutionary relationship of 47 of the 48 species (except bamboo shark) whose receptors were included in this study were depicted. The presence of receptor genes was marked by the filled circles, which are color-coded according to phylogenetic group as in fig. 1A. The emergence and loss of TARLL, TARL, and TAAR clades are indicated by colored stars and cross. Names of species in red text indicate a species whose receptors were cloned and tested in this study. The asterisks signify that ligands for TAARs of those species were not studied before. (B) The numbers of receptors cloned and deorphaned were summarized, and divided into 4 groups: receptors newly deorphaned in this study (red), receptors with novel ligands identified in this study (light green), receptors deorphaned in previous studies (light blue), and receptors cloned and tested but not deorphaned (grey).

**Figure 3. Ligand response profile of HTR4s, TARLs, TARLLs, and TAARs.** (A) Overview of the ligand response profile of all of the deorphaned receptors. The red squares represent positive ligand-receptor interaction and the yellow squares denote no responses. The horizontal lines show borders of different clades. The tested compounds were classified into 8 chemical clusters (Supplementary fig. S4) (only 6 clusters of compounds were found to be ligands of any receptor); different clusters are separated by the vertical lines. The molecule structures underneath panel A are chosen as representatives of each chemical cluster. The names of those ligands are shown below: 1: Butylamine; 2: Hexylamine; 3: Isoamylamine; 4: Isobutylamine; 5: Cadaverine; 6: Pyrrolidine; 7: Putrescine; 8: 2-methylbutylamine; 9: 3-methylthiopropylamine; 10: Agmatine; 11: Cyclohexylamine; 12: Cyclopentylamine; 13:
Piperidine; 14: Spermidine; 15: 5-amino-1-pentanol; 16: Spermine; 17: Amylamine; 18: Octylamine; 19: Heptylamine; 20: Propylamine; 21: Benzyamine; 22: Tryptamine; 23: Phenylethylamine; 24: Histamine; 25: Tyramine; 26: Octopamine; 27: 3-methoxytyramine; 28: 2-Phenylethylamine; 29: Histamine; 30: Tyramine; 31: Gramine; 32: Phenylethanolamine; 33: Dopamine; 34: Serotonin; 35: N,N-Dimethylcyclohexylamine; 36: 1-methylpyrrolidine; 37: 1-methylpiperidine; 38: Trimethylamine; 39: N,N-dimethylphenethylamine; 40: N,N-dimethyl benzyl amine; 41: N,N-dimethylisopropylamine; 42: N,N,N,N'-Tetramethyl-1,4-butanediamine; 43: N,N,N,N'-Tetramethyl-1,3-propanediamine; 44: Triethylamine; 45: Isopropylamine; 46: Aniline; 47: Indole; 48: 1-methylindole; 49: 3-methylindole; 50: 5-methylindole; 51: 6-methylindole; 52: 7-methylindole. The tuning breadth of each receptor is color-coded on the right. (B) The UMAP reduction and visualization of phyloPCA results of the deorphaned receptors is shown. The cluster in the lower right quadrant is outlined with a dashed rectangle and highlighted in the inset. Within this cluster, dots with black outlines indicate the receptors (HTR4, TARL, clade Ia TAAR, and a TARLL) which belong to group 2 by the PAM clustering method; dots are colored according to phylogenetic clade. Note that the receptor names of this group (PAM group 2) are colored in black in (A); the names of PAM group 1 receptors close to PAM group 2 are italicized in (A); all other names of PAM group 1 receptors are in grey.

Figure 4. Structural basis for ligand recognition of TARLLs. (A) Responses of sea lamprey TARLL4a to different groups of chemicals including aldehyde, ketone, acid, ester, alcohol, and amines (pink color). (B) Dose-dependent responses of TARLL4a to phenylethylamine and butylamine. (C) Dose-dependent responses of TARLL3h to N,N,N',N'-tetramethyl-1,4-butanediamine and N,N,N',N'-tetramethyl-1,3-propanediamine. (D) Homology modeling and ligand docking for TARLL3h binding to 2 ligands, N,N,N',N'-tetramethyl-1,4-butanediamine (left) and N,N,N',N'-tetramethyl-1,3-propanediamine (right). Both Asp3.32 and Tyr3.33 were found to interact with ligands. (E) Dose-dependent responses of TARLL1b to indole and 1-methylindole. (F) Homology modeling and ligand docking for TARLL3h binding to indole and 1-methylindole. Phe5.38 was found to play a role in ring recognition of indole through pi-pi interactions.

Figure 5. Ligand profiles and expression patterns of HTR4, TARL, and TAAR1. (A) qPCR analysis of different receptors in zebrafish tissues. Gapdh serves as a control. (B) Response of HTR4, TARL, TAARC1, and TAAR1 to 16 ligands at 10 μM. The heatmap is based on the ratio of fluorescence intensity of ligand-receptor interaction and no ligand controls in the luciferase assay. TAARC1 is also included because it resides in clade Ia close to TAAR1 and shows a similar ligand profile to HTR4, TARL, and TAAR. (C) PhyloPCA analysis on the
responses of HTR4, TARL, TAARC1, and TAAR1 to these 16 ligands. In this analysis, PC1 explains 70.9% of the variance and PC2 explains 19.5% of the variance. Dots colored by receptor clades represent the ligand responses of receptors on the first two PC axes. Ligand eigenvectors are represented by lines and colored according to the ligand clusters. A ligand name is shown at the tip of the line if its eigenvector has a higher contribution (compared to other ligands) to either PC1 or PC2. Red and blue dashed circles indicate two groups of receptors separated by PC1 and PC2. (D) The presence of HTR4, TARL1, TARL2, and TAAR1 in several representative species in vertebrates is plotted. (E) A model for the evolution and functional diversification of aminergic receptors. The clades labeled in green (including TARLL, TAARS1, TAARS2, the non-TAAR1 part of clade Ia, clade Ib, clade II, and clade III TAARs) play important roles in olfaction. The clades labeled in red, including HTR4, TARL, and TAAR1 are likely to detect amine neurotransmitters and trace amines in the brain.

**Supplementary Figure S1.** Phylogenetic tree showing the evolutionary history of HTR4, TARL, TARLL, and TAARs. (A-B) The phylogeny was inferred by maximum likelihood using IQ-TREE. The numbers on the nodes show the ultrafast bootstrap approximation (% of 5,000 total replicates, A) or the standard bootstrap (% of 100 total replicates, B). The 4 subfamilies, TARLL1-4 in the TARLL clade and the 2 subfamilies, TARL1 and TARL2 in the TARL clade are labeled in A.

**Supplementary figure S2.** Sequence conservation of receptors in different clades and pseudogene analysis. (A) Sequence logo plot of 4 major TAAR clades (clade Ia, Ib, II, and III), TARLL, TARL, and HTR4. The grey boxes indicate the 7 transmembrane segments predicted by GPCRdb. TM, transmembrane segments; ICL, intracellular loop; ECL, extracellular loop. (B) The numbers of TAAR/TARL/TARLL functional genes and pseudogenes as well as the percentages of pseudogenes were plotted.

**Supplementary figure S3.** Hierarchical clustering of molecules used for receptor deorphanization. The 97 amine molecules are classified into 8 clusters according to their structural similarities. A larger length of the line between 2 molecules indicates larger structural dissimilarity. The molecular structures in the right are the 52 identified ligands for HTR4, TARL, TARLL, and TAAR.

**Supplementary figure S4.** PhyloPCA results of all receptors and optimal number of clusters in PAM clustering algorithm. (A-C) The first 6 principal components (PC) of phyloPCA results, including PC1-2 (A), PC3-4 (B), and PC5-6 (C) are displayed. Dots colored by receptor clades represent the ligand responses of receptors in each new axis. Ligand
eigenvectors are shown by lines and colored by the ligand clusters. A ligand name is labeled at the tip of the line if its eigenvector has a higher weight for either PC in each plot. (D) The optimal number of receptor groups (here, 2) is determined and visualized by estimating the average silhouette width for PAM clustering using the phyloPCA score matrix. The x axis is the number of clusters measured. The y axis indicates how similar a data point within a cluster is to its own cluster compared to other clusters averaging across the number of clusters.

Supplementary figure S5. Comparison of different phylogenetic tree topologies and UMAP visualizations of the phyloPCA and PAM clustering results. (A-B) The tree trimmed from the phylogenetic tree in the study of Dieris, et al. 2021 (A) and in this study (B) are shown. (C-D) The UMAP visualization results of phyloPCA analyses on the trimmed tree in A (C) and B (D) are displayed. Similar to fig. 3B, we defined 2 groups using PAM clustering. The receptors represented by dots with black outlines belong to one group (PAM group 2 in fig. 3B), and other receptors belong to the other group.

Supplementary figure S6. Ligand profile and amine-recognition motifs of TAAR subfamily. (A) Sequence logo plot demonstrating conservation of 2 critical binding sites of TAARs, Asp$_{3.32}$ and Asp$_{5.42}$. The black arrow heads indicate the loss of one amino acid at 5.44 in clade II TAARs. (B) Dose-dependent responses of TAAR6 and TAAR8 to monoamines (N,N-dimethylcyclohexylamine) and diamines (cadaverine and putrescine). (C) The pattern of Asp$_{3.32}$ and Asp$_{5.42}$/Asp$_{5.43}$ in the phylogenetic tree. TAARs containing only Asp$_{3.32}$ (red), only Asp$_{5.42}$/Asp$_{5.43}$ (green, Asp$_{5.43}$ in some clade II TAAR members), both Asp$_{3.32}$ and Asp$_{5.42}$/Asp$_{5.43}$ (blue, Asp$_{3.32}$ and Asp$_{5.43}$ in some clade II TAAR members), or neither Asp$_{3.32}$ nor Asp$_{5.42}$ (black) are depicted.

Supplementary figure S7. Clade-specific ligand profiles and molecular mechanism of clade Ib and II TAARs. (A) Pie chart demonstrating the percentage of ligands in the chemical clusters recognized by clade Ib and II TAARs. (B) Dose-dependent responses and the corresponding EC$_{50}$ of some clade Ib II TAARs to their ligands. Clade Ib TAARs tend to recognize ligands with aromatic rings in chemical cluster 3 but clade II TAARs tend to recognize tertiary amines in chemical cluster 5.

Supplementary figure S8. Similarity of ligand profile for TAARS1 and TAAR5, and results of l1ou analyses. (A) Dose dependent responses of bamboo shark TAARS1b to 1-methylpiperidine, N,N-dimethylisopropylamine, and trimethylamine, which are ligands of TAAR5 as well. (B) Evolutionary shifts (asterisks and thick branches with different colors) in ligand profiles across receptor clades. Shifts were determined by l1ou on the first 5 PC scores.
using the pBIC criterion. Among 29 shifts detected, in addition to species-specific shifts, 2 evolutionary shifts in particular are worth noting (in yellow shading): a shift in TARL1, as well as in TAARS1. The bars on the right represent the receptor scores on each PC.

**Supplementary figure S9. Ligand specificity of TARLLs for amines.** (A) *In situ* hybridization of *Tarll1b*, *Tarll4a*, and *Tarll3i* genes in the sea lamprey MOE. Arrows indicate positive cells. (B) Dose-dependent response of TARLL4a to 8 primary monoamines from chemical cluster 2. (C) Site-directed mutagenesis of 2 amino acids predicted by computational docking. Mutation of Asp$^{3.32}$ and Tyr$^{3.33}$ into Ala (D3.32A and Y3.33A) completely eliminates the receptor activation. Mutation of Tyr$^{3.33}$ into Phe (Y3.33F) preserves part of the receptor function. (D) Responses of sea lamprey TARLL1b to different groups of chemicals including aldehyde, ketone, acid, ester, alcohol, and amines. (E) Dose-dependent response of sea lamprey TARLL1b for 6 indole derivatives, including tryptamine and serotonin.

**Supplementary figure S10. Comparison of l1ou results on trimmed trees.** l1ou analyses based on trimmed trees in the study of Dieris, et al. 2021 (A) and this study (B) were performed. The shifts in TARL1 (in yellow shading) were observed in both analyses. Evolutionary shifts (asterisks and thick branches with different colors) in ligand profiles across receptor clades. All the PCs were applied in l1ou analyses but only the results from the first 5 PCs are shown. The bars on the right represent the receptor scores on each PC.

**Supplementary figure S11. Dose-response curves of HTR4, TARL1, TARL2, and TAAR1 from different species to tryptamine and serotonin.** (A) Dose-dependent responses and the corresponding EC$_{50}$ of HTR4, TARL, TARL2, and TAAR1 to tryptamine. (B) Dose-dependent responses and the corresponding EC$_{50}$ of HTR4, TARL, TARL2, and TAAR1 to serotonin.

**Supplementary table S1. Summary of receptor numbers from each subclade chosen for phylogenetic tree construction.** The two sheets show the numbers of included receptors in the TARLL, TARL, TAAR subfamilies as well the HTR4s and outgroups.

**Supplementary table S2. Summary of cloned receptors.** The numbers of receptors cloned from different species in this study are summarized.

**Supplementary table S3 Summary of deorphaned receptors.** The responses of receptors deorphaned in this study are summarized, together with responses of receptors reported by previous studies.
Supplementary table S4. Summary of dose-dependent response curves. The dose-dependent response curves of deorphaned receptors and the corresponding EC₅₀ (µM) are summarized.

Supplementary table S5. Summary of zebrafish TAARs. TAARs reported by Hussian, et al. 2009 and Saraiva, et al. 2015 were aligned, edited, and renamed as indicated. We first confirmed the sequences from Saraiva et al., by searching in Ensembl database by the Ensembl IDs. We found 8 pseudogenes and 2 repeated sequences, and deleted these sequences. We then deleted abnormal exons and complemented their CDS regions according to the genome sequence. Next, we compared the sequences to those reported by Hussian et al. The results were divided into 4 groups: identity = 100% and coverage = 100%; identity = 100% and coverage < 100%; identity > 90%; no match. For the first 3 groups, we named them according to Hussian et al. For the last group, we named them according to their locations in the phylogenetic tree.

Supplementary table S6. List of chemicals and related information. Summary of chemicals used for cell-based high-throughput screening.

Supplementary table S7. List of primers for in situ probe production and qPCR. Summary of primers used for in situ probe production and qPCR.

Supplementary data S1. The amino acid sequences of receptors used for phylogenetic tree construction. All of the amino acid sequences in FASTA format used for phylogenetic tree construction were included.

Supplementary data S2. The nucleotide sequences of pseudogenes. The nucleotide sequences of identified TAAR/TARL/TARLL pseudogenes in alligator, coelacanth, spotted gar, whale shark, and sea lamprey were included.

References
Achat-Mendes C, Lynch LJ, Sullivan KA, Vallender EJ, Miller GM. 2012. Augmentation of methamphetamine-induced behaviors in transgenic mice lacking the trace amine-associated receptor 1. Pharmacol Biochem Behav 101:201-207.
Borowsky B, Adham N, Jones KA, Raddatz R, Artymyshyn R, Ogozalek KL, Durkin MM, Lakhani PP, Bonini JA, Pathirana S, et al. 2001. Trace amines: identification of a family of mammalian G protein-coupled receptors. Proc Natl Acad Sci U S A 98:8966-8971.

Bradaia A, Trube G, Stalder H, Norcross RD, Ozmen L, Wettstein JG, Pinard A, Buchy D, Gassmann M, Hoener MC, et al. 2009. The selective antagonist EPPTB reveals TAAR1-mediated regulatory mechanisms in dopaminergic neurons of the mesolimbic system. Proc Natl Acad Sci U S A 106:20081-20086.

Bunzow JR, Sonders MS, Arttamangkul S, Harrison LM, Zhang G, Quigley DI, Darland T, Suchland KL, Pasumamula S, Kennedy JL, et al. 2001. Amphetamine, 3,4-methylenedioxymethamphetamine, lysergic acid diethylamide, and metabolites of the catecholamine neurotransmitters are agonists of a rat trace amine receptor. Mol Pharmacol 60:1181-1188.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972-1973.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res 14:1188-1190.

Dewan A. 2020. Olfactory signaling via trace amine-associated receptors. Cell Tissue Res.

Dewan A, Cichy A, Zhang J, Miguel K, Feinstein P, Birnberg D, Bozza T. 2018. Single olfactory receptors set odor detection thresholds. Nat Commun 9:2887.

Dewan A, Pacifico R, Zhan R, Birnberg D, Bozza T. 2013. Non-redundant coding of aversive odours in the main olfactory pathway. Nature.

Dieris M, Kowatschew D, Korsching SI. 2021. Olfactory function in the trace amine-associated receptor family (TAARs) evolved twice independently. Sci Rep 11:7807.

Dietschi Q, Tuberosa J, Rosingh L, Loichot G, Ruedi M, Carleton A, Rodriguez I. 2017. Evolution of immune chemoreceptors into sensors of the outside world. Proc Natl Acad Sci U S A 114:7397-7402.

Durant JL, Leland BA, Henry DR, Nourse JG. 2002. Reoptimization of MDL Keys for Use in Drug Discovery. Journal of Chemical Information and Computer Sciences 42:1273-1280.
1 Eyun SI. 2019. Accelerated pseudogenization of trace amine-associated receptor genes in
2 primates. Genes Brain Behav 18:e12543.
3 Eyun SI, Moriyama H, Hoffmann FG, Moriyama EN. 2016. Molecular Evolution and
4 Functional Divergence of Trace Amine-Associated Receptors. PLoS One 11:e0151023.
5 Farid R, Day T, Friesner RA, Pearlstein RA. 2006. New insights about HERG blockade
6 obtained from protein modeling, potential energy mapping, and docking studies. Bioorg Med
7 Chem 14:3160-3173.
8 Fei A, Wu W, Tan L, Tang C, Xu Z, Huo X, Bao H, Kong Y, Johnson M, Hartmann G, et al.
9 2021. Coordination of two enhancers drives expression of olfactory trace amine-associated
10 receptors. Nat Commun 12:3798.
11 Ferrero DM, Lemon JK, Fluegge D, Pashkovski SL, Korzan WJ, Datta SR, Spehr M, Fendt
12 M, Liberles SD. 2011. Detection and avoidance of a carnivore odor by prey. Proc Natl Acad
13 Sci U S A.
14 Ferrero DM, Wacker D, Roque MA, Baldwin MW, Stevens RC, Liberles SD. 2012. Agonists
15 for 13 Trace Amine-Associated Receptors Provide Insight into the Molecular Basis of Odor
16 Selectivity. ACS Chem Biol.
17 Gainetdinov RR, Hoener MC, Berry MD. 2018. Trace Amines and Their Receptors.
18 Pharmacol Rev 70:549-620.
19 Gao S, Liu S, Yao J, Li N, Yuan Z, Zhou T, Li Q, Liu Z. 2017. Genomic organization and
20 evolution of olfactory receptors and trace amine-associated receptors in channel catfish,
21 Ictalurus punctatus. Biochim Biophys Acta Gen Subj 1861:644-651.
22 Greer PL, Bear DM, Lassance JM, Bloom ML, Tsukahara T, Pashkovski SL, Masuda FK,
23 Nowlan AC, Kirchner R, Hoekstra HE, et al. 2016. A Family of non-GPCR Chemosensors
24 Defines an Alternative Logic for Mammalian Olfaction. Cell 165:1734-1748.
25 Hashiguchi Y, Nishida M. 2007. Evolution of trace amine associated receptor (TAAR) gene
26 family in vertebrates: lineage-specific expansions and degradations of a second class of
27 vertebrate chemosensory receptors expressed in the olfactory epithelium. Mol Biol Evol
28 24:2099-2107.
Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol 35:518-522.

Hussain A, Saraiva LR, Ferrero DM, Ahuja G, Krishna VS, Liberles SD, Korsching SI. 2013. High-affinity olfactory receptor for the death-associated odor cadaverine. Proc Natl Acad Sci U S A 110:19579-19584.

Hussain A, Saraiva LR, Korsching SI. 2009. Positive Darwinian selection and the birth of an olfactory receptor clade in teleosts. Proc Natl Acad Sci U S A 106:4313-4318.

Izquierdo C, Gomez-Tamayo JC, Nebel JC, Pardo L, Gonzalez A. 2018. Identifying human diamine sensors for death related putrescine and cadaverine molecules. PLoS Comput Biol 14:e1005945.

Jia L, Li S, Dai W, Guo L, Xu Z, Scott AM, Zhang Z, Ren J, Zhang Q, Dexheimer TS, et al. 2021. Convergent olfactory trace amine-associated receptors detect biogenic polyamines with distinct motifs via a conserved binding site. J Biol Chem:101268.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587-589.

Kassambara A, Mundt F. 2020. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package v.1.07.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772-780.

Kelley LA, Gardner SP, Sutcliffe MJ. 1996. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. Protein Eng 9:1063-1065.

Khabbazian M, Kriebel R, Rohe K, Ané C. 2016. Fast and accurate detection of evolutionary shifts in Ornstein–Uhlenbeck models. Methods in Ecology and Evolution 7:811-824.

Li Q, Korzan WJ, Ferrero DM, Chang RB, Roy DS, Buchi M, Lemon JK, Kaur AW, Stowers L, Fendt M, et al. 2013. Synchronous evolution of an odor biosynthesis pathway and behavioral response. Curr Biol 23:11-20.
Li Q, Liberles SD. 2016. Odor sensing by trace amine-associated receptors. In: Zufall F, Munger SD, editors. Chemosensory transduction: the detection of odors, tastes, and other chemostimuli. Amsterdam; Boston: Elsevier/Academic Press. p. 67-80.

Li Q, Tachie-Baffour Y, Liu Z, Baldwin MW, Kruse AC, Liberles SD. 2015. Non-classical amine recognition evolved in a large clade of olfactory receptors. Elife 4:e10441.

Libants S, Carr K, Wu H, Teeter JH, Chung-Davidson YW, Zhang Z, Wilkerson C, Li W. 2009. The sea lamprey Petromyzon marinus genome reveals the early origin of several chemosensory receptor families in the vertebrate lineage. BMC Evol Biol 9:180.

Liberles SD, Buck LB. 2006. A second class of chemosensory receptors in the olfactory epithelium. Nature 442:645-650.

Liu J, Wu R, Li JX. 2020. TAAR1 and Psychostimulant Addiction. Cell Mol Neurobiol 40:229-238.

Maechler M, Rousseeuw P, Hubert M, Hornik K. 2019. Cluster: cluster analysis basics and extensions. R package v.2.1.0.

Maestro. 2020. schrödinger Release New York: Schrödinger, LLC.

Marra NJ, Stanhope MJ, Jue NK, Wang M, Sun Q, Pavinski Bitar P, Richards VP, Komissarov A, Rayko M, Klinger S, et al. 2019. White shark genome reveals ancient elasmobranch adaptations associated with wound healing and the maintenance of genome stability. Proc Natl Acad Sci U S A 116:4446-4455.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268-274.

Pacifico R, Dewan A, Cawley D, Guo C, Bozza T. 2012. An Olfactory Subsystem that Mediates High-Sensitivity Detection of Volatile Amines. Cell Rep 2:76-88.

Pandy-Szekeres G, Munk C, Tsonkov TM, Mordalski S, Harpsoe K, Hauser AS, Bojarski AJ, Gloriam DE. 2018. GPCRdb in 2018: adding GPCR structure models and ligands. Nucleic Acids Res 46:D440-D446.
Pfister P, Smith BC, Evans BJ, Brann JH, Trimmer C, Sheikh M, Arroyave R, Reddy G, Jeong HY, Raps DA, et al. 2020. Odorant Receptor Inhibition Is Fundamental to Odor Encoding. Curr Biol.

Revel FG, Moreau JL, Gainetdinov RR, Bradaia A, Sotnikova TD, Mory R, Durkin S, Zbinden KG, Norcross R, Meyer CA, et al. 2011. TAAR1 activation modulates monoaminergic neurotransmission, preventing hyperdopaminergic and hypoglutamatergic activity. Proc Natl Acad Sci U S A 108:8485-8490.

Revell LJ. 2009. Size-correction and principal components for interspecific comparative studies. Evolution 63:3258-3268.

Revell LJ, Collar DC. 2009. Phylogenetic analysis of the evolutionary correlation using likelihood. Evolution 63:1090-1100.

Saraiva LR, Ahuja G, Ivandic I, Syed AS, Marioni JC, Korsching SI, Logan DW. 2015. Molecular and neuronal homology between the olfactory systems of zebrafish and mouse. Sci Rep 5:11487.

Saraiva LR, Kondoh K, Ye X, Yoon KH, Hernandez M, Buck LB. 2016. Combinatorial effects of odorants on mouse behavior. Proc Natl Acad Sci U S A.

Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des 27:221-234.

Sastry M, Lowrie JF, Dixon SL, Sherman W. 2010. Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments. J Chem Inf Model 50:771-784.

Scott AM, Zhang Z, Jia L, Li K, Zhang Q, Dexheimer T, Ellsworth E, Ren J, Chung-Davidson YW, Zu Y, et al. 2019. Spermine in semen of male sea lamprey acts as a sex pheromone. PLoS Biol. 17:e3000332.

Shah A, Ratkowski M, Rosa A, Feinstein P, Bozza T. 2021. Olfactory expression of trace amine-associated receptors requires cooperative cis-acting enhancers. Nat Commun 12:3797.
Sharma K, Ahuja G, Hussain A, Balfanz S, Baumann A, Korsching SI. 2016. Elimination of a ligand gating site generates a supersensitive olfactory receptor. Sci Rep 6:28359.

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 11:398-405.

Tibshirani R. 1996. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological) 58:267-288.

Xu Z, Li Q. 2020. TAAR Agonists. Cell Mol Neurobiol 40:257-272.

Zhang D, Gao F, Jakovlic I, Zou H, Zhang J, Li WX, Wang GT. 2020. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour 20:348-355.

Zhang J, Pacifico R, Cawley D, Feinstein P, Bozza T. 2013. Ultrasensitive detection of amines by a trace amine-associated receptor. J Neurosci 33:3228-3239.

Zhang J, Yang J, Jang R, Zhang Y. 2015. GPCR-I-TASSER: A Hybrid Approach to G Protein-Coupled Receptor Structure Modeling and the Application to the Human Genome. Structure 23:1538-1549.
**Figure 2**

A. Phylogenetic tree showing the evolutionary relationships among different species. The tree branches are color-coded to indicate the emergence of different receptor types and the loss of certain receptor types from different lineages. Asterisks denote species included in the study.

B. Bar chart illustrating the number of receptors identified in different clades. The chart compares the receptors deorphaned in this study, those with novel ligands identified, and those deorphaned in previous studies. Each clade is represented by a set of bars, with the number of receptors indicated on the y-axis.
| Clade Ia                  | Clade Ib                  | Clade II              | Clade III             |
|-------------------------|---------------------------|-----------------------|-----------------------|
| Zebrafish HTR4          | Sea lamprey TARLL1b       | Sea lamprey TARLL3a   | Sea lamprey TARLL4a   |
| Mouse HTR4              | Shark TARL2               | Medaka TARL1          | Catfish TARL1         |
| Sea lamprey TARLL3b     | Zebrafish TARL1           | Catfish TARL1         | Fugu TARL1            |
| Sea lamprey TARLL5a     | Medaka TARL1              | Spotted gar TARL1     | Medaka TARL2          |
| Sea lamprey TARLL4a     | Coelacanth TARL1          | Coelacanth TARL1      | Coelacanth TARL2      |
| Sea lamprey TARLL6a     | Pollock TARL1             | Clade Ia              | Clade Ib              |
| Sea lamprey TARLL7a     | Clade II                  | Clade III             | Clade IV              |
| Sea lamprey TARLL9a     | Clade IV                  | Clade V               | Clade VI              |
| Sea lamprey TARLL10a    | Clade V                   | Clade VI              | Clade VII             |
| Sea lamprey TARLL12a    | Clade VII                 | Clade VIII             | Clade IX              |
| Sea lamprey TARLL14a    | Clade VIII                | Clade IX              | Clade X               |
| Sea lamprey TARLL16a    | Clade X                   | Clade X               | Clade XI              |
| Sea lamprey TARLL18a    | Clade XI                  | Clade XI              | Clade XII             |
| Sea lamprey TARLL20a    | Clade XII                 | Clade XII             | Clade XIII            |
| Sea lamprey TARLL22a    | Clade XIII                | Clade XIII            | Clade XIV             |
| Sea lamprey TARLL24a    | Clade XIV                 | Clade XIV             | Clade XV              |
| Sea lamprey TARLL26a    | Clade XV                  | Clade XV              | Clade XVI             |
| Sea lamprey TARLL28a    | Clade XVI                 | Clade XVI             | Clade XVII            |
| Sea lamprey TARLL30a    | Clade XVII                | Clade XVII            | Clade XVIII           |
| Sea lamprey TARLL32a    | Clade XVIII               | Clade XVIII           | Clade XIX             |
| Sea lamprey TARLL34a    | Clade XIX                 | Clade XIX             | Clade XX              |
| Sea lamprey TARLL36a    | Clade XX                  | Clade XX              | Clade XXI             |
| Sea lamprey TARLL38a    | Clade XXI                 | Clade XXI             | Clade XXII            |
| Sea lamprey TARLL40a    | Clade XXII                | Clade XXII            | Clade XXIII           |
| Sea lamprey TARLL42a    | Clade XXIII               | Clade XXIII           | Clade XXIV            |
| Sea lamprey TARLL44a    | Clade XXIV                | Clade XXIV            | Clade XXV             |
| Sea lamprey TARLL46a    | Clade XXV                 | Clade XXV             | Clade XXVI            |
| Sea lamprey TARLL48a    | Clade XXVI                | Clade XXVI            | Clade XXVII           |
| Sea lamprey TARLL50a    | Clade XXVII               | Clade XXVII           | Clade XXVIII          |
| Sea lamprey TARLL52a    | Clade XXVIII              | Clade XXVIII          | Clade XXIX            |
| Sea lamprey TARLL54a    | Clade XXIX                | Clade XXIX            | Clade XXX              |
| Sea lamprey TARLL56a    | Clade XXX                 | Clade XXX             | Clade XXXI             |
| Sea lamprey TARLL58a    | Clade XXXI                | Clade XXXI            | Clade XXXII            |
| Sea lamprey TARLL60a    | Clade XXXII               | Clade XXXII           | Clade XXXIII          |
| Sea lamprey TARLL62a    | Clade XXXIII              | Clade XXXIII          | Clade XXXIV            |
| Sea lamprey TARLL64a    | Clade XXXIV               | Clade XXXIV           | Clade XXXV             |
| Sea lamprey TARLL66a    | Clade XXXV                | Clade XXXV             | Clade XXXVI            |
| Sea lamprey TARLL68a    | Clade XXXVI               | Clade XXXVI           | Clade XXXVII          |
| Sea lamprey TARLL70a    | Clade XXXVII              | Clade XXXVII          | Clade XXXVIII         |
| Sea lamprey TARLL72a    | Clade XXXVIII             | Clade XXXVIII         | Clade XXXIX            |
| Sea lamprey TARLL74a    | Clade XXXIX               | Clade XXXIX           | Clade XXXX             |
| Sea lamprey TARLL76a    | Clade XXXX                | Clade XXXX           | Clade XXXI              |
| Sea lamprey TARLL78a    | Clade XXXI                | Clade XXXI            | Clade XXXII            |
| Sea lamprey TARLL80a    | Clade XXXII               | Clade XXXII           | Clade XXXIII          |
| Sea lamprey TARLL82a    | Clade XXXIII              | Clade XXXIII          | Clade XXXIV            |
| Sea lamprey TARLL84a    | Clade XXXIV               | Clade XXXIV           | Clade XXXV             |
| Sea lamprey TARLL86a    | Clade XXXV                | Clade XXXV             | Clade XXXVI            |
| Sea lamprey TARLL88a    | Clade XXXVI               | Clade XXXVI           | Clade XXXVII          |
| Sea lamprey TARLL90a    | Clade XXXVII              | Clade XXXVII          | Clade XXXVIII         |
| Sea lamprey TARLL92a    | Clade XXXVIII             | Clade XXXVIII         | Clade XXXIX            |
| Sea lamprey TARLL94a    | Clade XXXIX               | Clade XXXIX           | Clade XXXX             |
| Sea lamprey TARLL96a    | Clade XXXX                | Clade XXXX           | Clade XXXI              |
| Sea lamprey TARLL98a    | Clade XXXI                | Clade XXXI            | Clade XXXII            |
| Sea lamprey TARLL100a   | Clade XXXII               | Clade XXXII           | Clade XXXIII          |
| Sea lamprey TARLL102a   | Clade XXXIII              | Clade XXXIII          | Clade XXXIV            |
| Sea lamprey TARLL104a   | Clade XXXIV               | Clade XXXIV           | Clade XXXV             |
| Sea lamprey TARLL106a   | Clade XXXV                | Clade XXXV             | Clade XXXVI            |
| Sea lamprey TARLL108a   | Clade XXXVI               | Clade XXXVI           | Clade XXXVII          |
| Sea lamprey TARLL110a   | Clade XXXVII              | Clade XXXVII          | Clade XXXVIII         |
| Sea lamprey TARLL112a   | Clade XXXVIII             | Clade XXXVIII         | Clade XXXIX            |
| Sea lamprey TARLL114a   | Clade XXXIX               | Clade XXXIX           | Clade XXXX             |
| Sea lamprey TARLL116a   | Clade XXXX                | Clade XXXX           | Clade XXXI              |
| Sea lamprey TARLL118a   | Clade XXXI                | Clade XXXI            | Clade XXXII            |
| Sea lamprey TARLL120a   | Clade XXXII               | Clade XXXII           | Clade XXXIII          |
| Sea lamprey TARLL122a   | Clade XXXIII              | Clade XXXIII          | Clade XXXIV            |
| Sea lamprey TARLL124a   | Clade XXXIV               | Clade XXXIV           | Clade XXXV             |
| Sea lamprey TARLL126a   | Clade XXXV                | Clade XXXV             | Clade XXXVI            |
| Sea lamprey TARLL128a   | Clade XXXVI               | Clade XXXVI           | Clade XXXVII          |
| Sea lamprey TARLL130a   | Clade XXXVII              | Clade XXXVII          | Clade XXXVIII         |
| Sea lamprey TARLL132a   | Clade XXXVIII             | Clade XXXVIII         | Clade XXXIX            |
| Sea lamprey TARLL134a   | Clade XXXIX               | Clade XXXIX           | Clade XXXX             |
| Sea lamprey TARLL136a   | Clade XXXX                | Clade XXXX           | Clade XXXI              |
| Sea lamprey TARLL138a   | Clade XXXI                | Clade XXXI            | Clade XXXII            |
| Sea lamprey TARLL140a   | Clade XXXII               | Clade XXXII           | Clade XXXIII          |
| Sea lamprey TARLL142a   | Clade XXXIII              | Clade XXXIII          | Clade XXXIV            |
| Sea lamprey TARLL144a   | Clade XXXIV               | Clade XXXIV           | Clade XXXV             |
| Sea lamprey TARLL146a   | Clade XXXV                | Clade XXXV             | Clade XXXVI            |
| Sea lamprey TARLL148a   | Clade XXXVI               | Clade XXXVI           | Clade XXXVII          |
| Sea lamprey TARLL150a   | Clade XXXVII              | Clade XXXVII          | Clade XXXVIII         |
| Sea lamprey TARLL152a   | Clade XXXVIII             | Clade XXXVIII         | Clade XXXIX            |
| Sea lamprey TARLL154a   | Clade XXXIX               | Clade XXXIX           | Clade XXXX             |

**Figure 3**

A) Heatmap showing ligand clusters for different clades. B) UMAP projection of ligand clusters.
Figure 4
