

A vertebrate-wide catalogue of T1R receptors reveals diversity in taste perception

Received: 6 May 2023

Accepted: 25 October 2023

Published online: 13 December 2023

 Check for updates

Hidegori Nishihara^{1,2,9}✉, Yasuka Toda^{3,9}, Tae Kuramoto^{1,4}, Kota Kamohara³, Azusa Goto³, Kyoko Hoshino³, Shinji Okada⁵, Shigehiro Kuraku^{6,7}, Masataka Okabe⁸ & Yoshiro Ishimaru³✉

Taste is a vital chemical sense for feeding behaviour. In mammals, the umami and sweet taste receptors comprise three members of the taste receptor type 1 (T1R/*TASIR*) family: T1R1, T1R2 and T1R3. Because their functional homologues exist in teleosts, only three *TASIR* genes generated by gene duplication are believed to have been inherited from the common ancestor of bony vertebrates. Here, we report five previously uncharacterized *TASIR* members in vertebrates, *TASIR4*, *TASIR5*, *TASIR6*, *TASIR7* and *TASIR8*, based on genome-wide survey of diverse taxa. We show that mammalian and teleost fish *TASIR2* and *TASIR3* genes are paralogues. Our phylogenetic analysis suggests that the bony vertebrate ancestor had nine *TASIRs* resulting from multiple gene duplications. Some *TASIRs* were lost independently in descendent lineages resulting in retention of only three *TASIRs* in mammals and teleosts. Combining functional assays and expression analysis of non-teleost fishes we show that the novel T1Rs form heterodimers in taste-receptor cells and recognize a broad range of ligands such as essential amino acids, including branched-chain amino acids, which have not been previously considered as T1R ligands. This study reveals diversity of taste sensations in both modern vertebrates and their ancestors, which might have enabled vertebrates to adapt to diverse habitats on Earth.

Taste is one of the most important senses that govern the feeding behaviour of animals. It is widely accepted that mammals have five basic tastes: umami (savoury), sweet, bitter, salty and sour^{1,2}. Taste receptor type 1 (T1R, encoded by *TASIR*), a G protein-coupled receptor family, consists of three members, namely T1R1, T1R2 and T1R3, which are encoded by the genes *TASIR1*, *TASIR2* and *TASIR3*, respectively, and act as umami or sweet receptors^{3,4}. The T1R1/T1R3 heterodimer functions as an umami taste receptor in mammals and detects L-amino acids and 5'-ribonucleotides⁵⁻⁷. The mammalian T1R2/T1R3 heterodimer acts as a sweet sensor^{6,8}. Likewise, homologues of *TASIR* family genes have been

identified in teleost fishes⁹, and each of the heterodimers T1R1/T1R3 and T1R2/T1R3 can sense several amino acids in teleosts¹⁰.

A previous phylogenetic analysis revealed that all mammalian and teleost *TASIRs* can be grouped into the *TASIR1*, *TASIR2* and *TASIR3* clades¹¹, suggesting that their common ancestor had only three T1R members derived from gene duplications that have been retained in present-day species. Lineage-specific duplications and losses of *TASIR* genes have occurred within each of the *TASIR1*, *TASIR2* and *TASIR3* clades, as exemplified by multiple *TASIR2* genes in zebrafish and fugu, and loss of *TASIR2* in birds¹². A few genomic studies of vertebrates such

¹Department of Advanced Bioscience, Graduate School of Agriculture, Kindai University, Nara, Japan. ²School of Life Science and Technology, Tokyo Institute of Technology, Yokohama, Japan. ³Department of Agricultural Chemistry, School of Agriculture, Meiji University, Kawasaki, Japan. ⁴Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Japan. ⁵Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan. ⁶Molecular Life History Laboratory, National Institute of Genetics, Mishima, Japan. ⁷Department of Genetics, SOKENDAI (Graduate University for Advanced Studies), Mishima, Japan. ⁸Department of Anatomy, The Jikei University School of Medicine, Tokyo, Japan. ⁹These authors contributed equally: Hidegori Nishihara, Yasuka Toda. ✉e-mail: nishihara@nara.kindai.ac.jp; yishimaru@meiji.ac.jp

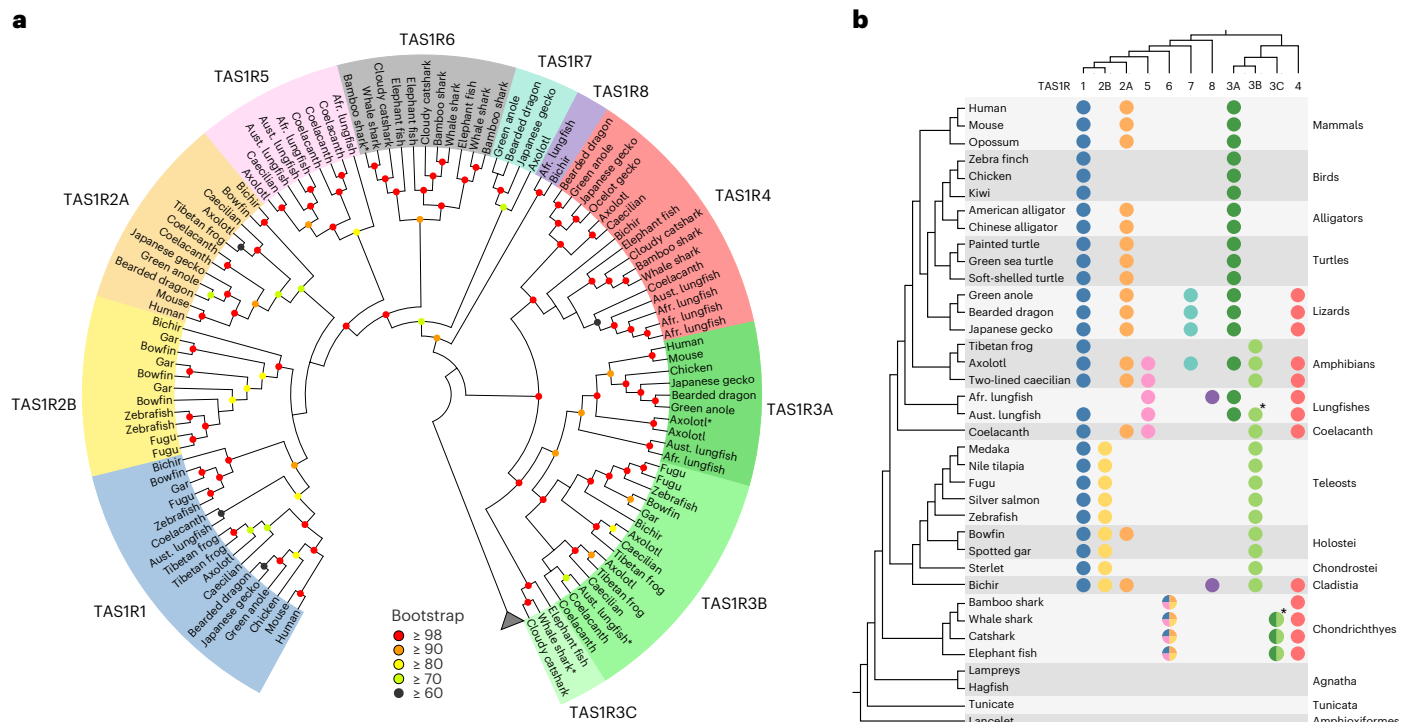


Fig. 1 | Phylogenetic tree and the revised classification of *TASIR* members.
a, Maximum-likelihood tree for amino acid sequences inferred from *TASIR*s for 21 jawed vertebrates constructed with the JTT + G (CAT approximation) model in RAxML. Coloured circles in each node represent bootstrap values calculated with 1,000 replications, whereas those with low bootstrap support (<60) have no circles. Species classification is represented with coloured highlighting at the tips of the tree. GPRC6A was used as an outgroup (not shown). Afr, African; Aust, Australian. **b**, Distribution of *TASIR* members among chordates. The

colour of circles corresponds to the coloured highlighting in **a** and indicates the presence of *TASIR* members in the genome assemblies of the various chordates. Phylogenetic relationships among species and among *TASIR*s are shown on the left and top, respectively. *TASIR6* of cartilaginous fishes is the orthologue of the *TASIR1/2A/2B/5* clade and is shown as a circle with assorted colours. Similarly, *TASIR3C* of cartilaginous fishes is shown with two shades of green that represent *TASIR3A* and *TASIR3B*. Circles with asterisks denote putative pseudogenes.

as squamates, coelacanth and sharks have suggested the existence of taxonomically unplaced *TASIR*s that may not be included in the aforementioned three clades^{13–15}. However, the lack of comprehensive characterization and systematic classification has limited our understanding of the evolutionary history of *TASIR* genes, the functional diversity of TIRs, and the molecular basis of taste sense in vertebrates.

Here, we present an evolutionary analysis of diverse *TASIR*s in jawed vertebrates, with an exhaustive taxon sampling encompassing all major ‘fish’ lineages. In addition to clades *TASIR1*, *TASIR2* and *TASIR3*, we identified five novel *TASIR* clades. The results suggest that the vertebrate ancestor possessed more TIRs than most modern vertebrates, challenging the paradigm that only three TIR family members have been retained during evolution. Functional analyses suggest that the novel TIRs have shaped the diversity of taste sense. We propose that the TIR family has undergone an ancient birth-and-death evolution that accelerated their functional differentiation, which may have led to the diversification of feeding habitats among vertebrates.

Results

Identification of novel *TASIR* family members

We identified homologues of *TASIR* genes that are included in public genome/transcriptome databases for diverse taxa of jawed vertebrates (Supplementary Table 1). Except for jawed vertebrates, *TASIR* genes were not identified in any Deuterostomia reference genomes (lampreys, hagfishes, tunicates, lancelets, sea urchins, starfish, hemichordate, etc.) or the nr database, suggesting that the *TASIR*/TIR family exists only in jawed vertebrates. All phylogenetic trees, as estimated using different methods and datasets, consistently revealed the existence of many *TASIR*s that had not been categorized into any of the three known clades: *TASIR1*,

TASIR2 and *TASIR3*. These previously undocumented *TASIR*s were found in lizards, amphibians, lungfishes, coelacanth, bichir and cartilaginous fishes (Fig. 1a and Extended Data Figs. 1–3). The novel *TASIR*s could be classified into five new clades. One clade, which is the sister clade of *TASIR3*, was named *TASIR4* and contains genes from all jawed vertebrates investigated except mammals, birds, crocodylians, turtles, frog, sterlets or neopterygians (Fig. 1b and Extended Data Fig. 4). Another novel *TASIR*, named *TASIR5*, exists in axolotl, lungfishes and coelacanth and is close to the clade comprising *TASIR1* and *TASIR2* (Fig. 1a).

The sister clade to *TASIR1* + *TASIR2* + *TASIR5*, which was named *TASIR6*, was identified exclusively in cartilaginous fishes. *TASIR6* could be further divided into three subclades, namely *TASIR6-1*, *TASIR6-2* and *TASIR6-3*, all of which were found to be present in elephant fish (also called elephant shark), belonging to the taxon Holocephali of cartilaginous fishes (Extended Data Figs. 1–3). Therefore, the three *TASIR6* subclades probably emerged in the common ancestor of extant cartilaginous fishes. A thorough search of the genomes and transcriptomes of the four cartilaginous fish species identified only *TASIR3*, *TASIR4* and *TASIR6*, but no orthologues of *TASIR1*, *TASIR2* or *TASIR5* (Fig. 1b and Extended Data Fig. 4), suggesting that the *TASIR1*, *TASIR2* and *TASIR5* genes in bony vertebrates are co-orthologues of the *TASIR6* genes in cartilaginous fish.

Another novel *TASIR* clade, *TASIR7*, was found exclusively in axolotl and lizards. Yet another new clade, *TASIR8*, was identified only in bichir and lungfishes, and its monophyly was robustly supported (Fig. 1 and Extended Data Figs. 1–3), suggesting that *TASIR8* emerged in the common ancestor of bichir and lungfishes. Indeed, the likelihood of an alternative relationship, in which *TASIR7* and *TASIR8* form an exclusive cluster and represent a species tree, was rejected statistically based on the approximately unbiased test ($P < 10^{-4}$; Extended Data Fig. 5),

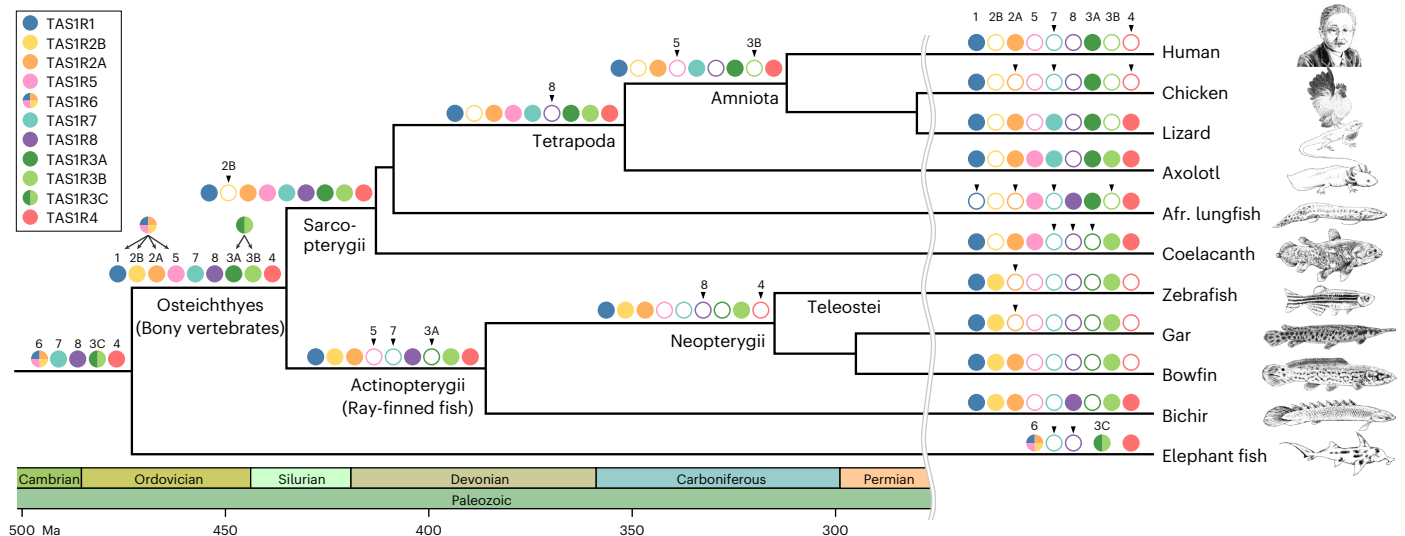


Fig. 2 | Birth-and-death history of the *TASIR* family genes during vertebrate evolution. The colour key indicates the names of the various *TASIR* members. Filled, coloured circles on the branches indicate the presence of *TASIR* members, whereas open circles indicate their absence, as estimated based on the phylogenetic tree (Fig. 1a) and distribution among vertebrates (Fig. 1a).

Arrowheads above open circles indicate that the *TASIR* member was lost at the branch. Geological periods and ages (Ma) taken from TimeTree¹⁶ are shown at the bottom. Taxon names are shown below branches. Species-specific gene duplication events for each *TASIR* were ignored. Illustrations of the species, including humans (Kikunae Ikeda, the discoverer of umami), are shown on the right.

suggesting that *TASIR7* and *TASIR8* are distinct groups. Among the vertebrates we investigated, the axolotl was found to possess *TASIRs* from the greatest number (seven) of clades (Fig. 1b and Supplementary Table 2).

Each of *TASIR3* and *TASIR2* consists of two paralogous clades

Remarkably, the phylogenetic analysis also revealed that *TASIR3* of bony vertebrates could be divided into two clades, named *TASIR3A* and *TASIR3B*, with high branch support (Fig. 1 and Extended Data Figs. 1–3). *TASIR3A* was found to be present in tetrapods and lungfishes but not other vertebrates, whereas *TASIR3B* was identified only in amphibians, lungfishes, coelacanth and ray-finned fishes. The sister clade to *TASIR3A* + *TASIR3B* was identified exclusively in cartilaginous fishes and named *TASIR3C*. This distribution suggested that an ancestral *TASIR3* gene was duplicated in the common ancestor of bony vertebrates, with subsequent independent loss of *TASIR3A* in certain lineages such as coelacanth and ray-finned fishes, whereas *TASIR3B* was lost in Amniota (mammals and sauropsids). Therefore, the *TASIR3* genes in mammals and teleost fishes are paralogues. Axolotl and Australian lungfish retained both *TASIR3A* and *TASIR3B* although the lungfish *TASIR3B* has been pseudogenized. Furthermore, the amphibians possess two groups of *TASIR3B*, named *TASIR3B1* and *TASIR3B2* (Fig. 1a), suggesting that *TASIR3B* was again duplicated—at the latest—before the common ancestor of amphibians.

A distinguishing feature of *TASIR3B* in ray-finned fishes is the presence of additional introns. In contrast to other *TASIRs*, which consist of six exons, exon 3 of *TASIR3B* in ray-finned fishes has been altered during evolution such that it now comprises two exons, suggesting the acquisition of an intron in the common ancestor of ray-finned fishes (Extended Data Fig. 6). Furthermore, exon 6 of *TASIR3B* in non-bichir ray-finned fishes acquired an additional intron, resulting in a total of eight exons of the gene. Thus, this intron is likely to have been inserted after the divergence of bichir. Except for these two instances, the exon-intron structure is conserved among the *TASIR* genes we investigated.

Also, *TASIR2* does not form a single clade in the tree (Fig. 1). The *TASIR2* genes in ray-finned fishes form a clade with *TASIR1*, and the other *TASIR2* group from tetrapods, lungfish, coelacanth, bowfin and bichir forms a sister group to the clade comprising *TASIR1* and the ray-finned fish *TASIR2*. The paraphyletic relationship of the two *TASIR2* groups is concordant with previous reports¹⁵. Hereafter, we refer to the major

vertebrate group as *TASIR2A* and the ray-finned fish group as *TASIR2B* (Fig. 1). Notably, we found that the anciently diverged ray-finned fishes such as bowfin and bichir retained both *TASIR2A* and *TASIR2B* as well as *TASIR1*. We assessed the likelihood of other phylogenetic relationships in which *TASIR2s* have a single origin, and the hypotheses were significantly rejected ($P < 10^{-6}$, approximately unbiased test; Extended Data Fig. 5). These results suggested that the *TASIR2* genes in mammals and teleost fishes are paralogues. Thus, the *TASIR* phylogenetic tree comprised a total of 11 *TASIR* clades: *TASIR1*, *TASIR2A*, *TASIR2B*, *TASIR3A*, *TASIR3B*, *TASIR3C*, *TASIR4*, *TASIR5*, *TASIR6*, *TASIR7* and *TASIR8*. This unexpected gene diversity challenges conventional conceptions about the evolution of the genetic basis for umami and sweet receptors.

Birth-and-death evolution of the *TASIR* family

Some of the higher-level relationships among the *TASIR* clades were supported with relatively high branch support, as exemplified by the exclusive cluster of *TASIR3* + *TASIR4*, the clade of the other *TASIRs*, the clade of *TASIR1* + *TASIR2B* + *TASIR2A* + *TASIR5*, and the sister relationship of this latter clade to *TASIR6* (Fig. 1). Based on the phylogenetic relationships and the distribution of all *TASIR* members (Fig. 1b), the most parsimonious evolutionary scenario could be deduced as follows (Fig. 2). The first *TASIR* gene emerged in the ancestral lineage of jawed vertebrates during the period 615–473 million years ago (Ma) according to TimeTree¹⁶. This ancestral *TASIR* underwent multiple duplications to produce at least five *TASIR* genes: *TASIR3C* (the ancestral gene of *TASIR3A* and *TASIR3B*), *TASIR4*, *TASIR7*, *TASIR8* and *TASIR6* (the ancestral gene of *TASIR1*, *TASIR2B*, *TASIR2A* and *TASIR5*). Owing to speciation between cartilaginous fishes and bony vertebrates ~473 Ma, *TASIR6* and the ancestral gene of clade *TASIR1* + *TASIR2B* + *TASIR2A* + *TASIR5* diverged. This speciation probably also led to the split between *TASIR3C* and clade *TASIR3A* + *TASIR3B*. In the stem lineage of bony vertebrates (473–435 Ma), *TASIR1*, *TASIR2A*, *TASIR2B* and *TASIR5* were generated via additional gene duplication events. Simultaneously, *TASIR3A* and *TASIR3B* were generated by gene duplication, resulting in a total of nine *TASIRs* in the common ancestor of bony vertebrates (Fig. 2). After the divergence of ray-finned and lobe-finned fishes ~435 Ma, a portion of the expanded *TASIRs* began to be differentially lost during vertebrate evolution. For example, *TASIR8* was lost in the tetrapod ancestor, *TASIR3B* and *TASIR5* were lost in the amniote ancestor, and *TASIR4*

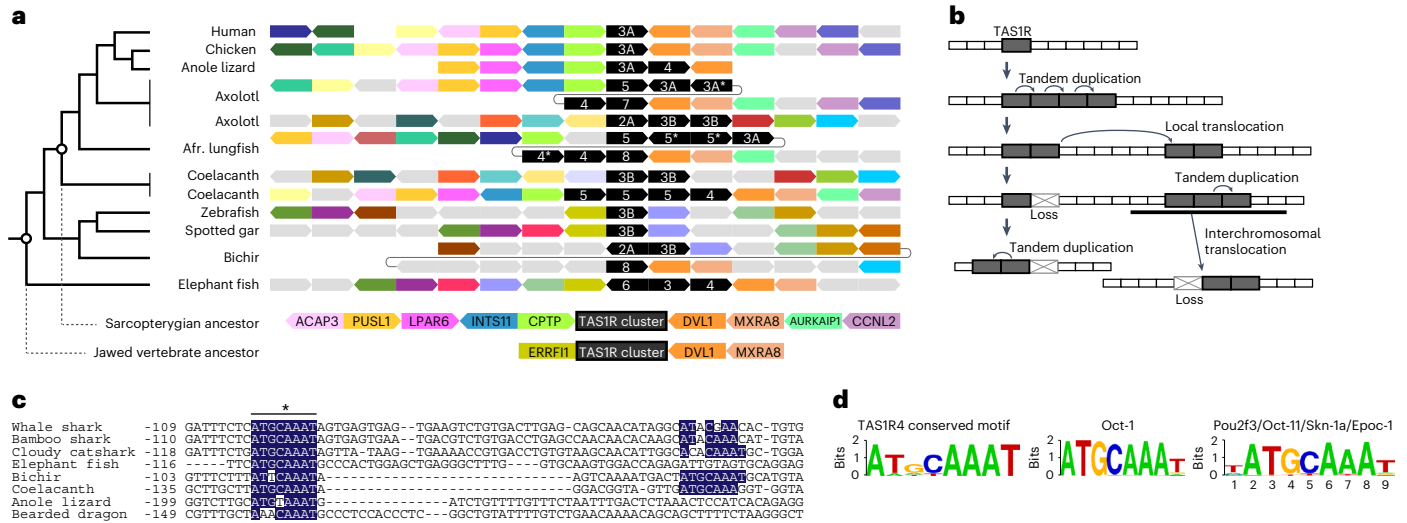


Fig. 3 | Synteny around *TASIRs* and conserved Oct-like motifs in the *TASIR4* upstream regions across vertebrates. **a, Synteny around each *TASIR* gene cluster is partly conserved across representative vertebrates. *TASIRs* are represented by black polygons, and those with asterisks are putative pseudogenes. Coloured polygons indicate genes shared among species, and grey colour represents genes not shared among the species or unknown. The species tree is shown on the left. The deduced gene orders in common ancestors of Sarcopterygii and jawed vertebrates are shown at the bottom. **b**, Proposed model for the expansion of *TASIR* genes across distant chromosomal regions during**

evolution. **c**, Conserved motifs located upstream of *TASIR4*. Sequence alignment of the upstream region of the *TASIR4* open reading frame revealed two conserved Oct-like transcription-factor binding motifs (blue shading). Numbers represent nucleotide positions from the *TASIR4* start codon site. The asterisk indicates one of the motifs that significantly resembles the Oct factor binding motif. **d**, Sequence logo for the conserved motif denoted with the asterisk in **c**. Known binding motifs of Oct-1 (retrieved from TRANSFAC) and Oct-11/Pou2f3/Skn-1a/Epoc-1 (retrieved from JASPAR) are compared.

and *TASIR7* were lost in the mammalian ancestor (Fig. 2). Thus, gene expansion before the common ancestor of bony vertebrates as well as the subsequent loss of a subset of genes have resulted in the rather dispersed distribution of *TASIRs* in extant species (Fig. 1b).

TASIR gene cluster revealed by scanning understudied genomes

The simplest model for gene amplification is a tandem duplication that produces multiple genes located side-by-side^{17,18}. However, *TASIR1*, *TASIR2* and *TASIR3* are located far from each other in both mammalian and teleost genomes. In human chromosome 1, for example, *TASIR1* is 12 Mb distant from *TASIR2A* and 5 Mb distant from *TASIR3A*, with many intervening genes in each case. In zebrafish, each of *TASIR1* and *TASIR3B* is located on a different chromosome from the two copies of *TASIR2B*, prompting us to hypothesize that *TASIR* members may have undergone expansion by tandem duplications in the ancestral genome, followed by subsequent translocation to distant regions during evolution. To address this possibility, the synteny of *TASIR3* and *TASIR4* was investigated among vertebrates, particularly those having the novel *TASIRs* (Fig. 3 and Extended Data Fig. 7). Indeed, the novel *TASIRs* were found to be located side-by-side in anole lizard, axolotl, lungfish, coelacanth and elephant fish (Fig. 3a). Even *TASIR2A* and *TASIR3B* are located next to each other in axolotl and bichir. This result suggested that a *TASIR* gene cluster had formed in the common ancestor of jawed vertebrates.

A comparison of neighbouring genes revealed that the *TASIR* cluster is flanked by two genes, namely *DVL1* and *MXRA8*, in the genomes of human, chicken, axolotl, lungfish, coelacanth, bichir and elephant fish (Fig. 3a), suggesting that these two genes were adjacent to the *TASIR* cluster in the common ancestor of jawed vertebrates. On the opposite end of the *TASIR* cluster, the gene order of *ACAP3–PUS11–LPAR6–INTS11–CPTP* may have been established in the sarcopterygian ancestor based on conservation among coelacanth, axolotl, chicken and partly in lizard. Furthermore, the presence of other *TASIR*-proximal genes is also conserved even across distant chromosomal regions (Extended Data Fig. 7). This suggested that a chromosomal region containing

both *TASIR* and multiple neighbouring genes—rather than the *TASIR* gene alone—had translocated to a different region in each lineage. Based on the inferred ancestral gene order, the unique distribution of *TASIRs* among present-day mammals and teleost fishes may have been a consequence of a combination of several events (Fig. 3b): (1) tandem duplication that produced a *TASIR* cluster in the ancestor of jawed vertebrates; (2) local translocation of a subset of *TASIRs* within a chromosome, as seen in multiple clusters observed in axolotl and coelacanth (Extended Data Fig. 7); (3) translocation of entire *TASIR*-containing regions to different chromosomes, as observed in zebrafish; and (4) gene loss(es) in each lineage, as partly observed as the presence of pseudogenes (Fig. 1a). Moreover, lineage-specific duplication events have occurred such as *TASIR2B* in zebrafish and *fugu* and *TASIR2A* in coelacanth (Fig. 1a and Extended Data Fig. 7)^{12,13}. Finally, we found that some of the *TASIRs* identified have been pseudogenized; for example, the whale shark *TASIR3C* and the lungfish *TASIR3B* (Fig. 1). These observations also support the evolutionary model of the *TASIR* family presented in Fig. 3b.

Conservation of a possible Oct-binding site in *TASIR4*

Because *TASIR4* is shared among a wide variety of vertebrates in contrast to the other novel *TASIRs*, we expected that a transcriptional regulatory mechanism might be conserved among the species. To explore the existence of a possible regulatory element, sequences upstream of the open reading frames of *TASIR4* from various species were aligned, and MEME¹⁹ was used to search for transcription-factor binding motifs conserved among the species. The most significant hit was the binding motif for the Oct family ($P < 10^{-12}$ for Oct-4, $P < 10^{-7}$ for Oct-1). At least one sequence of the known Oct-binding motif ‘ATGCAAAT’ is conserved among cartilaginous fishes, coelacanth, bichir and lizards in the region upstream of *TASIR4* (Fig. 3c,d). Although little is known about the transcriptional regulatory network in taste-receptor cells (TRCs), one known transcription factor responsible for TRC differentiation is Skn-1a, which is an Oct factor also known as Oct-11, Epoc-1 or Pou2f3 (ref. 20). In mammals, Skn-1a is exclusively expressed in umami, sweet and bitter TRCs, and loss of Skn-1a results

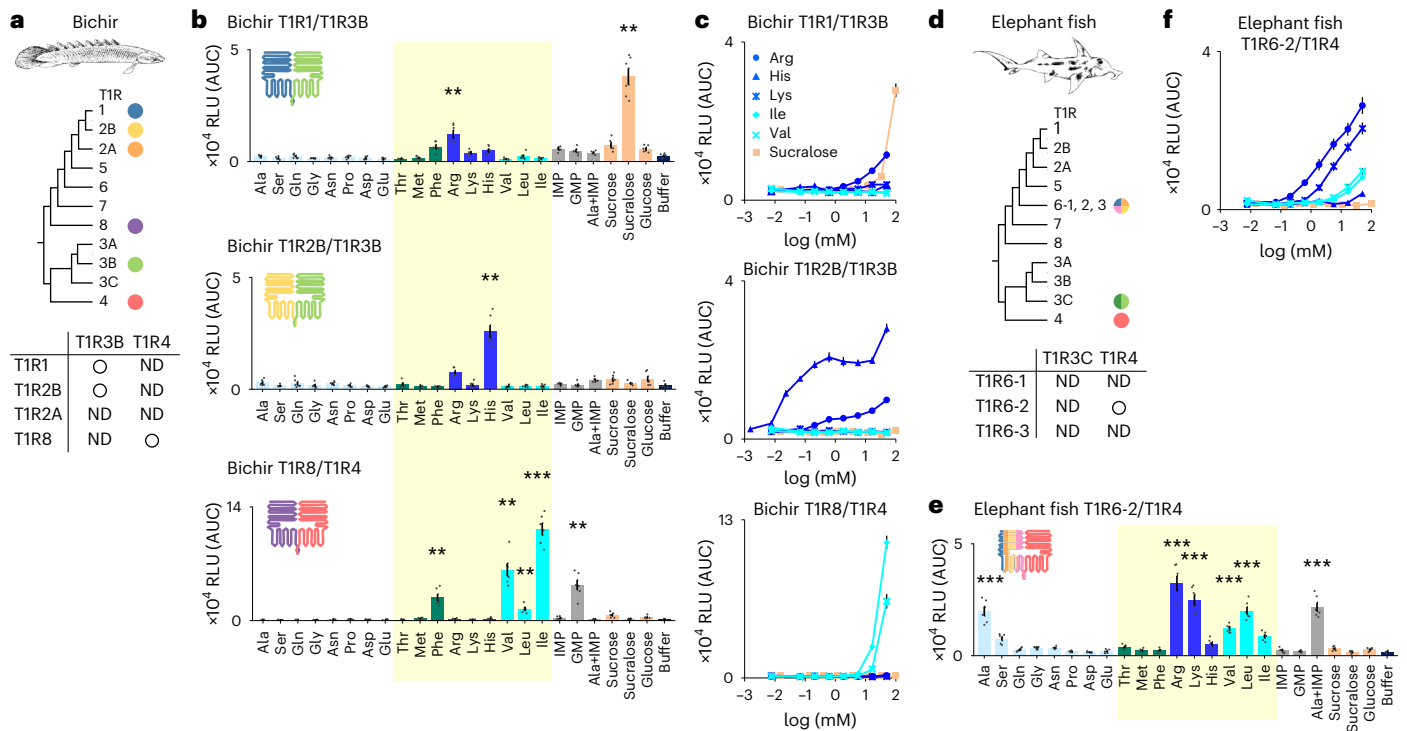


Fig. 4 | Functional analysis of T1Rs from bichir and elephant fish. a, T1R repertoire in bichir and their combinations used for the functional analysis. ND, not detected for any ligands tested. **b**, Responses of three combinations of T1R1/T1R3B (upper), T1R2B/T1R3B (middle) and T1R8/T1R4 (lower) to each of 17 amino acids (50 mM), nucleic acids (10 mM), sugars and sucralose (100 mM). Values represent the mean \pm s.e.m. of six independent experiments performed with duplicate samples. **, >10,000 RLU with $q < 0.01$; ***, >10,000 RLU with $q < 0.001$ by one-sided *t*-test (T1R1/T1R3B Arg, $P = 0.0012$; sucralose, $P = 0.00094$; T1R2B/T1R3B His, $P = 0.00015$; T1R8/T1R4 Phe, $P = 0.00063$; Val, $P = 0.00030$; Leu, $P = 0.00057$; Ile, $P = 0.000013$; GMP, $P = 0.00047$).

Amino acids that are essential in fishes are highlighted in yellow. AUC, area under the curve. **c**, Dose-response curves for T1R1/T1R3B (upper), T1R2B/T1R3B (middle) and T1R8/T1R4 (lower) to three basic amino acids (Arg, His and Lys; blue), two BCAAs (Ile and Val; light blue) and an artificial sweetener (sucralose; orange). Values represent the mean \pm s.e.m. of six independent experiments performed with duplicate samples. **d-f**, Same as **a-c**, respectively, for elephant fish and the functional analysis of T1R6-2/T1R4 (Ala, $P = 0.00015$; Arg, $P = 0.00013$; Lys, $P = 0.000094$; Val, $P = 0.000011$; Leu, $P = 0.000045$; Ala + IMP, $P = 0.000069$).

in the complete absence of these TRCs^{20,21}. This finding suggested that *TASIR4* expression is governed by a conserved regulatory mechanism involving an Oct transcription factor, possibly Skn-1a. Although Oct-binding sites were not observed in the other novel *TASIRs*, these findings may help to elucidate the molecular mechanisms underlying the conserved and/or lineage-specific expression of a variety of *TASIRs* in TRCs, which will enhance our understanding of the evolutionary origin of TRCs.

T1R diversity expands the range of taste sensation

Because receptor responses cannot be predicted from sequence analysis alone, functional tests using cultured cells heterologously expressing the target receptor are useful. We previously established a high-throughput screening system for the T1R receptors using a luminescence-based assay²² and have used it to identify ligands for both mammalian^{7,23} and non-mammalian²⁴⁻²⁶ T1R receptors. To examine which T1R receptors can form heterodimers and which ligands they respond to, we performed the functional analysis for the T1Rs of bichir, which possesses two newly discovered T1R groups (T1R4 and T1R8) and four known T1R groups (T1R1, T1R2A, T1R2B and T1R3B). Because *TASIR4* is the sister clade of *TASIR3* and is present in all vertebrates that harbour the other novel *TASIRs* (Fig. 1b), T1R4 could be assumed to form a heterodimer with another T1R. We combined either T1R3B or T1R4 with another T1R (T1R1, T1R2A, T1R2B or T1R8) in the functional analysis (Fig. 4a). Among these receptor pairs, strong responses to amino acids were detected for T1R1/T1R3B, T1R2B/T1R3B and T1R8/T1R4 (Fig. 4b and Extended Data Fig. 8). For bichir T1R2A,

its combination with T1R3B or T1R4 did not yield a response to any of the tastants examined (Extended Data Fig. 8a). Responses were not observed when T1R4 or T1R8 alone was used (Extended Data Fig. 8a), suggesting that these newly discovered T1Rs function as obligate heterodimers in bichir.

The bichir T1R8/T1R4 responded strongly to Phe and to branched-chain amino acids (BCAA; Ile, Val and Leu), whereas T1R1/T1R3B and T1R2B/T1R3B responded strongly to basic amino acids (Arg and His) (Fig. 4b,c). Fishes have 12 nutritionally essential amino acids (Cys, His, Ile, Leu, Lys, Met, Phe, Arg, Thr, Trp, Tyr and Val)²⁷, 9 of which are included in the 17 amino acids that were tested in the T1R functional analysis. Notably, all six amino acids to which the bichir T1Rs responded are essential amino acids ($P < 0.05$; one-sided Fisher's exact test), suggesting that the bichir T1Rs may sense essential amino acids in foods by taking advantage of the ability to perceive BCAA via the T1R4-related receptor.

Bichir T1R1/T1R3B also responded to sucralose, a structural analogue of sucrose. Although only T1R2A/T1R3A is responsible for sugar perception in mammals and lizards²⁶, we previously demonstrated that T1R1/T1R3A of birds has gained the ability to detect sugars^{24,25}. Also, T1R2B/T1R3B of two teleost fishes, namely carp²⁸ and gilthead seabream²⁹, can detect sugars at high concentrations (100–200 mM). Our assay was unable to analyse sugars at concentrations greater than 100 mM because of non-specific responses caused by changes in osmolarity. Although the sucrose response at 100 mM was not significantly higher than the thresholds we set in this study (>10,000 relative light units (RLU) with a false discovery rate (q) of <0.01), combined with

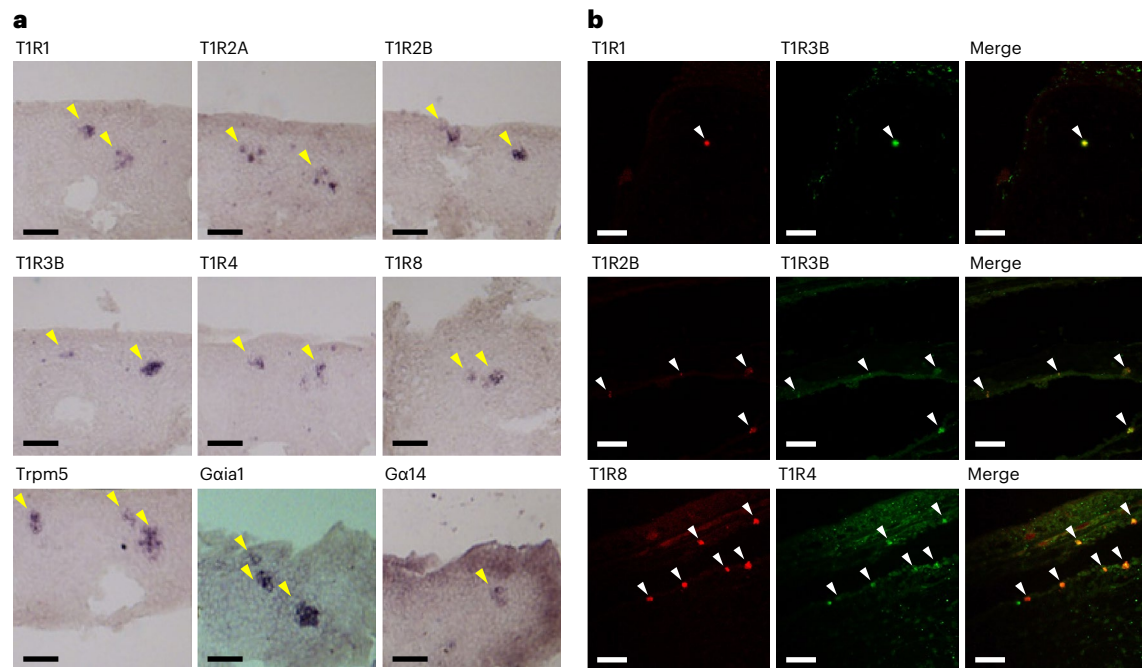


Fig. 5 | In situ hybridization of T1Rs in TRCs of bichir. a, Expression of six T1Rs and three marker genes in sagittal sections of the lips. Yellow arrowheads indicate TRCs that expressed the various genes. Scale bar, 50 μ m. The experiments were repeated at least three times. **b**, Double-label fluorescence in situ hybridization

for the combinations of T1R1/T1R3B (upper), T1R2B/T1R3B (middle) and T1R8/T1R4 (lower) in the sections. White arrowheads indicate co-expressing cells. Scale bar, 50 μ m. The experiments were repeated at least twice.

the fact that its structural analogue, sucralose, could elicit a clear response, higher concentrations of sucrose may be able to activate bichir T1R1/T1R3B. In addition, we found that bichir T1R8/T1R4 could respond to GMP, although a previous study reported that neither T1R1/T1R3B nor T1R2B/T1R3B of medaka fish nor T1R2B/T1R3B of zebrafish could be activated by 5'-ribonucleotides¹⁰. Therefore, the origin and evolution of sugar and nucleotide taste perception may need to be reconsidered based on results from future genetic and functional analyses of T1Rs.

We also performed a functional analysis of elephant fish T1Rs. Three genes of the T1R6 clade, namely T1R6-1, T1R6-2 and T1R6-3, were tested in combination with T1R3C and T1R4, and only the response of the T1R6-2/T1R4 pair could be detected (Fig. 4d–f and Extended Data Fig. 8b). This combination responded to a relatively broad range of amino acids, including both BCAA (Val, Leu) and basic amino acids (Arg, Lys). The T1Rs of mammals and teleosts have little or no response to BCAA but can respond to basic amino acids^{5,10,23}. The observed strong response of bichir T1R8/T1R4 and elephant fish T1R6-2/T1R4 to BCAA may reflect functional characteristics of the novel T1Rs involving T1R4 and possibly that of ancient T1Rs in the vertebrate ancestor.

Expression of the novel T1Rs in TRCs

To investigate whether the novel T1Rs are indeed expressed in TRCs, we performed in situ hybridization with sections of the lips and gill rakers of bichir (Fig. 5a). T1R1, T1R2A, T1R2B, T1R3B, T1R4 and T1R8 were expressed in subsets of TRCs. Genes encoding downstream signal-transduction molecules, such as TRPM5, $\text{G}\alpha\text{1}$ and $\text{G}\alpha\text{14}$, were also highly expressed in subsets of TRCs in the lips and gill rakers. The signal frequencies for TRPM5, $\text{G}\alpha\text{1}$ and $\text{G}\alpha\text{14}$ were higher than those for T1Rs.

To examine the localization of T1Rs in TRCs, we next performed double-label fluorescence in situ hybridization. This analysis confirmed the overlap of the signal for T1R1 with that of T1R3B, T1R2B with T1R3B and T1R8 with T1R4 (Fig. 5b). These results suggested that T1R1/T1R3B, T1R2B/T1R3B and T1R8/T1R4 function as heterodimers, in accordance with the results of our functional assays.

Discussion

The complex evolutionary history of the T1R/TAS1R family includes ancient gene expansions followed by independent lineage-specific losses, which contrasts with conventional wisdom that essentially only three members were retained during evolution^{11,30}. The evolution of certain other chemoreceptors, such as the T2R (or TAS2R) bitter-taste receptor family and olfactory receptors, followed a birth-and-death process³¹. In this mode of evolution, tens or hundreds of the receptor family/superfamily genes have undergone extensive lineage-specific duplication followed by frequent gene loss by deletion/inactivation³⁰. Our results suggest that a similar process—although less extensive than what occurred for other chemoreceptors—contributed to the phylogenetic and functional expansion of the T1R family early during vertebrate evolution. *TAS1Rs* were not subjected to extensive birth-and-death evolution, possibly because T1R ligands are limited to amino acids, sugars and nucleotides in contrast to T2Rs and olfactory receptors that respond to a wider range of ligands/stimulants. In line with our discovery, many chemoreceptors, including *TAS1Rs* in teleost fishes, have recently been reported to have undergone dynamic evolution including lineage-specific expansion and gene losses³². It is also possible that the ancient expansion might have contributed to an alternate use of T1Rs in tissues other than the sensory organs because certain G protein-coupled receptors (including T1Rs) are expressed in the gut of mammals and teleost fishes^{33,34} although their functions remain unresolved.

The functional combinations of the bichir T1R8/T1R4 and the elephant fish T1R6-2/T1R4 suggest that T1R4 may have a similar role to T1R3 by forming a functional heterodimer with another novel T1R such as T1R5, T1R6, T1R7 or T1R8. This model is supported by the fact that species with either *TAS1R5*, *TAS1R6*, *TAS1R7* or *TAS1R8* also have *TAS1R4* (Fig. 1b) and that *TAS1R4* is phylogenetically the sister group of *TAS1R3* (Fig. 1a). Therefore, the common ancestor of bony vertebrates, which had at least nine T1Rs, probably had two types of heterodimeric T1R receptors, namely T1R3- and T1R4-dependent receptors. This relatively wide variety of possible T1R combinations involving two duplicated

genes of T1R2 (A and B) and T1R3 (A and B) might have contributed to the diversification of taste sensation.

Our findings provoke new questions, one of which is why many *TASIR* genes—particularly the T1R4-related receptors—have become unnecessary in each lineage independently, and many species have come to rely predominantly on T1R3-dependent receptors (Fig. 2). One possible explanation is that dietary changes could have rendered one or more T1Rs unnecessary, and therefore, gene loss might have had little or no effect on survival. This is plausible because previous studies reported losses of *TASIRs* and *TAS2Rs* in many land vertebrates, presumably in association with specific dietary shifts^{32,35–37}. Also, the behaviour of swallowing foods whole, without mastication, could have diminished the essentiality of taste sense in certain vertebrates, as previously discussed with respect to mammals^{36,38} and squamates³⁹. Alternatively, it is possible that T1R3-dependent receptors have acquired greater functional flexibility and/or evolvability than other T1Rs; various tastants might have been detected via the evolutionary tuning of sequences and structures of the T1R3-dependent receptors rather than additional gene duplication. Such cases are indeed known for land vertebrates such as primates⁷ and birds^{24,25}. To address these issues, it will be essential to carry out functional analyses of the newly discovered T1Rs in addition to the known T1R1/T1R3 and T1R2/T1R3 for a broad range of vertebrates, as our current results demonstrate. For example, the response to BCAA is a previously unreported characteristic shared between the bichir T1R8/T1R4 and elephant fish T1R6-2/T1R4 (Fig. 4). This type of result provides insight into the sensory characteristics of an ancestor of vertebrates. We also found that bichir T1Rs responded to other essential amino acids, a sucrose analogue and a nucleotide. Future analysis will resolve whether the functions indeed reflect the characteristics of the ancestral species.

Thus, by demonstrating the unexpected diversity and unique evolutionary process of the T1R family, our results set the stage for understanding the evolutionary-scale changes in taste sense in vertebrates. The remarkably broad range of tastants detected by the T1Rs reflects the latent diversity of taste senses in vertebrates, and this may explain their successful expansion across diverse feeding habitats on Earth. Our understanding of taste sense will be further enhanced by clarifying T1R repertoires in each species, their tissue-specific expression, transcriptional regulatory mechanisms and protein structures. Revealing the functional and structural diversity of the novel T1Rs will also help us elucidate the molecular mechanisms by which human T1Rs recognize palatable tastes.

Methods

Identification of *TASIR* genes from genome and RNA sequencing data

We used genome and transcriptome data as well as related raw sequence reads for a broad range of vertebrates (Supplementary Table 1). First, a tblastn search was conducted against the 33 genomes using amino acid sequences of exon 6 of the *TASIRs* of human, chicken and zebrafish as queries. Hit sequences meeting the E value threshold of $1,110^{-40}$ were used to construct a phylogenetic tree using RAxML v.8.2.12 with the JTT + G (CAT approximation) model. The G protein-coupled receptor family C group 6 member A (GPC6A) genes, which are the closest relative of T1Rs⁴⁰, were used as the outgroup. Identified valid *TASIR* sequences were used for subsequent iterations of the tblastn search. RNA sequencing data were assembled using Bridger v.r2014-12-01 with default parameters and were used as a database for the tblastn search⁴¹.

We also conducted a NCBI tblastn search against all reference genomes of Deuterostomia excluding jawed vertebrates (Gnathostomata), and did not find any *TASIR* orthologues. In addition, an NCBI blastp search against the nr database, excluding Gnathostomata, yielded no hits for *TASIR* orthologues. Subsequently, we performed comprehensive annotation of *TASIR* exons in 21 organisms, including

model organisms and species that were presumed to possess novel/unclassified *TASIR* members, as identified via the procedure above.

The exon regions were predicted using AUGUSTUS v.3.2.3 (ref. 42), followed by an evaluation of the exon–intron boundaries by aligning the genome sequences with the human and zebrafish *TASIR* sequences and by the GT/AG rule. Because a certain degree of base errors was observed in the genome assembly for axolotl, sequence correction was needed for our *TASIR* identification. We retrieved the raw reads of the public genome data and RNA sequencing data corresponding to the *TASIR* exons using bowtie2 (ref. 43) and blastn and used that data to correct the *TASIR* sequences by checking the alignment. The *TASIR* amino acid sequences identified for axolotl, coelacanth and bichir were used as queries for an additional tblastn search of other vertebrates.

Phylogenetic analysis

For the full-length amino acid sequences, non-homologous residues were masked using PQUAL⁴⁴ and the sequences were aligned using MAFFT v.7.427 with the ginsi option⁴⁵. The phylogenetic tree was constructed using RAxML as described above. In addition, a maximum-likelihood tree was constructed under the posterior mean site frequency approximation⁴⁶ of the JTT + C20 + F + Γ model with 1,000 bootstrap replicates using IQ-TREE v.2.2.2.6 (ref. 47). Bayesian tree inference was conducted with MrBayes 3.2.6 with the JTT-F + Γ_4 model⁴⁸. Two simultaneous runs were carried out with 10,000,000 generations, of which 2,500,000 were discarded as burn-in, and convergence was assessed with Tracer⁴⁹. Trees were visualized with iTOL⁵⁰. Alternative tree topologies were evaluated with the approximately unbiased test with 100,000 replicates using CONSEL v.0.20 (ref. 51).

Syntenic analysis

The synteny of genes proximal to the novel T1Rs was analysed using annotations available in Ensembl 97 (ref. 52) for human (GRCh38), chicken (GRCg6a), anole lizard (AnoCar2.0), coelacanth (LatCha1), zebrafish (GRCz11) and spotted gar (LepOcu1). For bichir, annotations were conducted using Cufflinks on a draft assembly. The gene annotation for axolotl was obtained from the Axolotl-omics website (AmexG_v6.0-DD)⁵³. NCBI annotation was referred to for the West African lungfish (PAN1.0) and elephant fish (Callorhynchus_milii-6.1.3). Novel *TASIRs* were added to the gene list in our synteny analysis if they were not accurately identified in the public annotation data.

Conserved motifs in the sequence upstream of *TASIR4*

Sequences up to 300 bp upstream of the *TASIR4* open reading frames were collected for whale shark, bamboo shark, cloudy catshark, elephant fish, bichir, coelacanth, axolotl, two-lined caecilian, Japanese gecko, anole lizard and central bearded dragon. The sequences were aligned using MAFFT⁴⁵ and then used for MEME analysis¹⁹ to search for a maximum of three conserved sequence motifs. The motifs discovered by MEME were then used for comparison with known transcription-factor binding motifs in TRANSFAC v.11.3 using STAMP⁵⁴. The known Oct-11/Pou2f3 motif was obtained from JASPAR⁵⁵.

Experimental animals

This study was carried out in accordance with the National Institutes of Health guide for the care and use of laboratory animals (NIH Publication No. 8023, revised 1978). Both male and female bichir (*Polypterus senegalus*), ~5–7 cm body length, were purchased from a local commercial source. We found no differences in the expression of genes encoding T1Rs or downstream signal-transduction molecules, such as TRPM5, G α 1 and G α 14, between male and female bichir by in situ hybridization.

Cloning *TASIRs* of bichir and elephant fish

TASIR1, *TASIR2A*, *TASIR2B*, *TASIR3B*, *TASIR4* and *TASIR8* were amplified by PCR from the genomic DNA or cDNA of bichir. *TASIR6-1*, *TASIR6-2*, *TASIR6-3*, *TASIR3C* and *TASIR4* were amplified by PCR from the genomic

DNA of elephant fish (*Callorhynchus milii*). PCR and Sanger sequencing for the coding sequences of their *TASIR* genes were performed using specific primers designed based on the annotation from the whole genome assemblies. The PCR products of the exons were assembled into one full-length sequence using overlapping PCR (In-fusion cloning; Clontech) for each *TASIR* and were then subcloned into the pEAK10 expression vector (Edge Biosystems).

Functional analysis of T1Rs

Responses of the T1Rs to various taste-associated stimulants were measured using a cell-based luminescence assay, as described previously^{22,23}. Briefly, HEK293T cells were transiently co-transfected with an expression vector for an individual T1R along with a chimeric rat G protein (rG15i2) and a calcium-binding photoprotein (mt-apoclytin-II). Cells were seeded in 96-well plates and assayed 2 days after transfection. Cells were exposed to each taste stimulant individually, and luminescence intensity was measured using a FlexStation 3 microplate reader (Molecular Devices). The response in each well was calculated based on the area under the curve and expressed as RLU. Data were collected from three independent experiments, each carried out with duplicate samples. We adapted a strict definition for the positive response as >10,000 RLU along with a statistically significant difference against control (buffer) with a false discovery rate (*q*) of <0.01 (one-sided *t*-test). A limitation of this assay is that concentrations of amino acids and sugars were presented at a maximum of 50 mM or 100 mM to avoid receptor-independent calcium increases, caused for instance by changes in osmolarity²³, which can prevent the accurate assessment of responses to higher ligand concentrations. The osmotic pressure of each of the Arg and His solutions was higher than those of the other amino acid solutions because large amounts of HCl or NaOH were required for pH adjustment; this may have caused the higher response to 50 mM His of bichir T1R2B/T1R3B (Fig. 4c).

In situ hybridization

In situ hybridization was performed as described previously⁹. In brief, fresh-frozen sections (10 μm thick) of bichir jaw tissue were placed on MAS-coated glass slides (Matsunami Glass) and fixed with 4% paraformaldehyde in phosphate-buffered saline. Prehybridization (58 °C, 1 h), hybridization (58 °C, two overnight sessions), washing (58 °C, 0.2× saline–sodium citrate) and development (nitroblue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate; NBT-BCIP) were performed using digoxigenin-labelled probes. Images of stained sections were obtained using a fluorescence microscope (DM6 B; Leica) equipped with a cooled CCD digital camera (DFC7000 T; Leica). Double-label fluorescence in situ hybridization was performed using digoxigenin- and fluorescein-labelled RNA probes. Each labelled probe was detected sequentially by incubation with a peroxidase-conjugated antibody against digoxigenin and peroxidase-conjugated anti-fluorescein (Roche) followed by incubation with tyramide signal amplification (TSA)–Alexa Fluor 555 and TSA–Alexa Fluor 488 (Invitrogen) using the tyramide signal amplification method. Images of stained sections were obtained using a confocal laser-scanning microscope (LSM 800; ZEISS). The entire coding regions for the six T1Rs and two G protein α subunits as well as the partial coding region for Trpm5, all of which were amplified from bichir cDNA synthesized from lip tissue, were used as probes for in situ hybridization.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The *TASIR* sequences and phylogenetic trees are provided in Supplementary Data 1 and 2, respectively.

Code availability

No code was generated in this study.

References

- Trivedi, B. P. Gustatory system: the finer points of taste. *Nature* **486**, S2–S3 (2012).
- Yarmolinsky, D. A., Zuker, C. S. & Ryba, N. J. Common sense about taste: from mammals to insects. *Cell* **139**, 234–244 (2009).
- Li, X. et al. Human receptors for sweet and umami taste. *Proc. Natl Acad. Sci. USA* **99**, 4692–4696 (2002).
- Hummel, T. & Welge-Lüssen, A. *Taste and Smell: An Update* (Karger, 2006).
- Nelson, G. et al. An amino-acid taste receptor. *Nature* **416**, 199–202 (2002).
- Zhao, G. Q. et al. The receptors for mammalian sweet and umami taste. *Cell* **115**, 255–266 (2003).
- Toda, Y. et al. Evolution of the primate glutamate taste sensor from a nucleotide sensor. *Curr. Biol.* **31**, 4641–4649.e5 (2021).
- Nelson, G. et al. Mammalian sweet taste receptors. *Cell* **106**, 381–390 (2001).
- Ishimaru, Y. et al. Two families of candidate taste receptors in fishes. *Mech. Dev.* **122**, 1310–1321 (2005).
- Oike, H. et al. Characterization of ligands for fish taste receptors. *J. Neurosci.* **27**, 5584–5592 (2007).
- Shi, P. & Zhang, J. Contrasting modes of evolution between vertebrate sweet/umami receptor genes and bitter receptor genes. *Mol. Biol. Evol.* **23**, 292–300 (2006).
- Bachmanov, A. A. et al. Genetics of taste receptors. *Curr. Pharm. Des.* **20**, 2669–2683 (2014).
- Picone, B. et al. Taste and odorant receptors of the coelacanth—a gene repertoire in transition. *J. Exp. Zool. B* **322**, 403–414 (2014).
- Hara, Y. et al. Madagascar ground gecko genome analysis characterizes asymmetric fates of duplicated genes. *BMC Biol.* **16**, 40 (2018).
- Sharma, K., Syed, A. S., Ferrando, S., Mazan, S. & Korsching, S. I. The chemosensory receptor repertoire of a true shark is dominated by a single olfactory receptor family. *Genome Biol. Evol.* **11**, 398–405 (2019).
- Kumar, S. et al. TimeTree 5: an expanded resource for species divergence times. *Mol. Biol. Evol.* **39**, masc174 (2022).
- Ohno, S. *Evolution by Gene Editing* (Springer, 1970).
- Lewis, E. B. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570 (1978).
- Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
- Matsumoto, I., Ohmoto, M., Narukawa, M., Yoshihara, Y. & Abe, K. Skn-1a (Pou2f3) specifies taste receptor cell lineage. *Nat. Neurosci.* **14**, 685–687 (2011).
- Yamashita, J., Ohmoto, M., Yamaguchi, T., Matsumoto, I. & Hirota, J. Skn-1a/Pou2f3 functions as a master regulator to generate Trpm5-expressing chemosensory cells in mice. *PLoS ONE* **12**, e0189340 (2017).
- Toda, Y., Okada, S. & Misaka, T. Establishment of a new cell-based assay to measure the activity of sweeteners in fluorescent food extracts. *J. Agric. Food Chem.* **59**, 12131–12138 (2011).
- Toda, Y. et al. Two distinct determinants of ligand specificity in T1R1/T1R3 (the umami taste receptor). *J. Biol. Chem.* **288**, 36863–36877 (2013).
- Baldwin, M. W. et al. Sensory biology. Evolution of sweet taste perception in hummingbirds by transformation of the ancestral umami receptor. *Science* **345**, 929–933 (2014).
- Toda, Y. et al. Early origin of sweet perception in the songbird radiation. *Science* **373**, 226–231 (2021).

26. Liang, Q. et al. T1R2-mediated sweet sensing in a lizard. *Curr. Biol.* **32**, R1302–R1303 (2022).
27. Hou, Y. & Wu, G. Nutritionally essential amino acids. *Adv. Nutr.* **9**, 849–851 (2018).
28. Yuan, X. C. et al. Expansion of sweet taste receptor genes in grass carp (*Ctenopharyngodon idellus*) coincided with vegetarian adaptation. *BMC Evol. Biol.* **20**, 25 (2020).
29. Angotzi, A. R., Puchol, S., Cerda-Reverter, J. M. & Morais, S. Insights into the function and evolution of taste 1 receptor gene family in the carnivore fish gilthead seabream (*Sparus aurata*). *Int. J. Mol. Sci.* **21**, 7732 (2020).
30. Nei, M., Niimura, Y. & Nozawa, M. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat. Rev. Genet.* **9**, 951–963 (2008).
31. Nei, M. & Rooney, A. P. Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* **39**, 121–152 (2005).
32. Policarpo, M., Baldwin, M., Casane, D. & Salzburger, W. Diversity and evolution of the vertebrate chemoreceptor gene repertoire. Preprint at <https://doi.org/10.21203/rs.3.rs-2922188/v1> (2023).
33. Jang, H. J. et al. Gut-expressed gustducin and taste receptors regulate secretion of glucagon-like peptide-1. *Proc. Natl Acad. Sci. USA* **104**, 15069–15074 (2007).
34. Calo, J. et al. First evidence for the presence of amino acid sensing mechanisms in the fish gastrointestinal tract. *Sci. Rep.* **11**, 4933 (2021).
35. Antinucci, M. & Risso, D. A matter of taste: lineage-specific loss of function of taste receptor genes in vertebrates. *Front. Mol. Biosci.* **4**, 81 (2017).
36. Jiang, P. et al. Major taste loss in carnivorous mammals. *Proc. Natl Acad. Sci. USA* **109**, 4956–4961 (2012).
37. Liu, G. et al. Differentiated adaptive evolution, episodic relaxation of selective constraints, and pseudogenization of umami and sweet taste genes *TAS1Rs* in catarrhine primates. *Front. Zool.* **11**, 79 (2014).
38. Feng, P., Zheng, J., Rossiter, S. J., Wang, D. & Zhao, H. Massive losses of taste receptor genes in toothed and baleen whales. *Genome Biol. Evol.* **6**, 1254–1265 (2014).
39. Feng, P. & Liang, S. Molecular evolution of umami/sweet taste receptor genes in reptiles. *PeerJ* **6**, e5570 (2018).
40. Kuang, D. et al. Ancestral reconstruction of the ligand-binding pocket of Family C G protein-coupled receptors. *Proc. Natl Acad. Sci. USA* **103**, 14050–14055 (2006).
41. Chang, Z. et al. Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. *Genome Biol.* **16**, 30 (2015).
42. Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465–W467 (2005).
43. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
44. Whelan, S., Irisarri, I. & Burki, F. PREQUAL: detecting non-homologous characters in sets of unaligned homologous sequences. *Bioinformatics* **34**, 3929–3930 (2018).
45. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
46. Wang, H. C., Minh, B. Q., Susko, E. & Roger, A. J. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst. Biol.* **67**, 216–235 (2018).
47. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
48. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
49. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
50. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
51. Shimodaira, H. & Hasegawa, M. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**, 1246–1247 (2001).
52. Cunningham, F. et al. Ensembl 2022. *Nucleic Acids Res.* **50**, D988–D995 (2022).
53. Schloissnig, S. et al. The giant axolotl genome uncovers the evolution, scaling, and transcriptional control of complex gene loci. *Proc. Natl Acad. Sci. USA* **118**, e2017176118 (2021).
54. Mahony, S. & Benos, P. V. STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic Acids Res.* **35**, W253–W258 (2007).
55. Castro-Mondragon, J. A. et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **50**, D165–D173 (2022).

Acknowledgements

We thank S. Hyodo (The University of Tokyo) for providing the *Callorhynchus milii* sample. We also thank E. Kamiya (School of Life Science and Technology, Tokyo Institute of Technology) for technical assistance. The authors acknowledge Open Facility Center, Tokyo Institute of Technology, for sequencing assistance. Computations were partially performed on the supercomputer systems at the Research Organization of Information and Systems National Institute of Genetics and the Institute of Statistical Mathematics. This study was supported by Japan Society for the Promotion of Science KAKENHI grant nos. 19H03272 (to H.N.), 18K14427, 20H02941 and 23H02168 (to Y.T.), Research Project Grant(B) from the Institute of Science and Technology, Meiji University (to Y.I.), and the Lotte Shigemitsu Prize (to Y.T. and Y.I.).

Author contributions

H.N., Y.T. and Y.I. conceived and supervised the study. H.N., T.K., S.K. and M.O. analysed the vertebrate genomes. H.N. performed the phylogenetic and synteny analyses. Y.T. performed the functional assay. K.K., A.G., K.H., S.O. and Y.I. performed in situ hybridization experiments. H.N., Y.T. and Y.I. wrote the original draft of the manuscript. H.N., Y.T., Y.I., S.K. and M.O. edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41559-023-02258-8>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41559-023-02258-8>.

Correspondence and requests for materials should be addressed to Hidenori Nishihara or Yoshiro Ishimaru.

Peer review information *Nature Ecology & Evolution* thanks Iker Irisarri and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

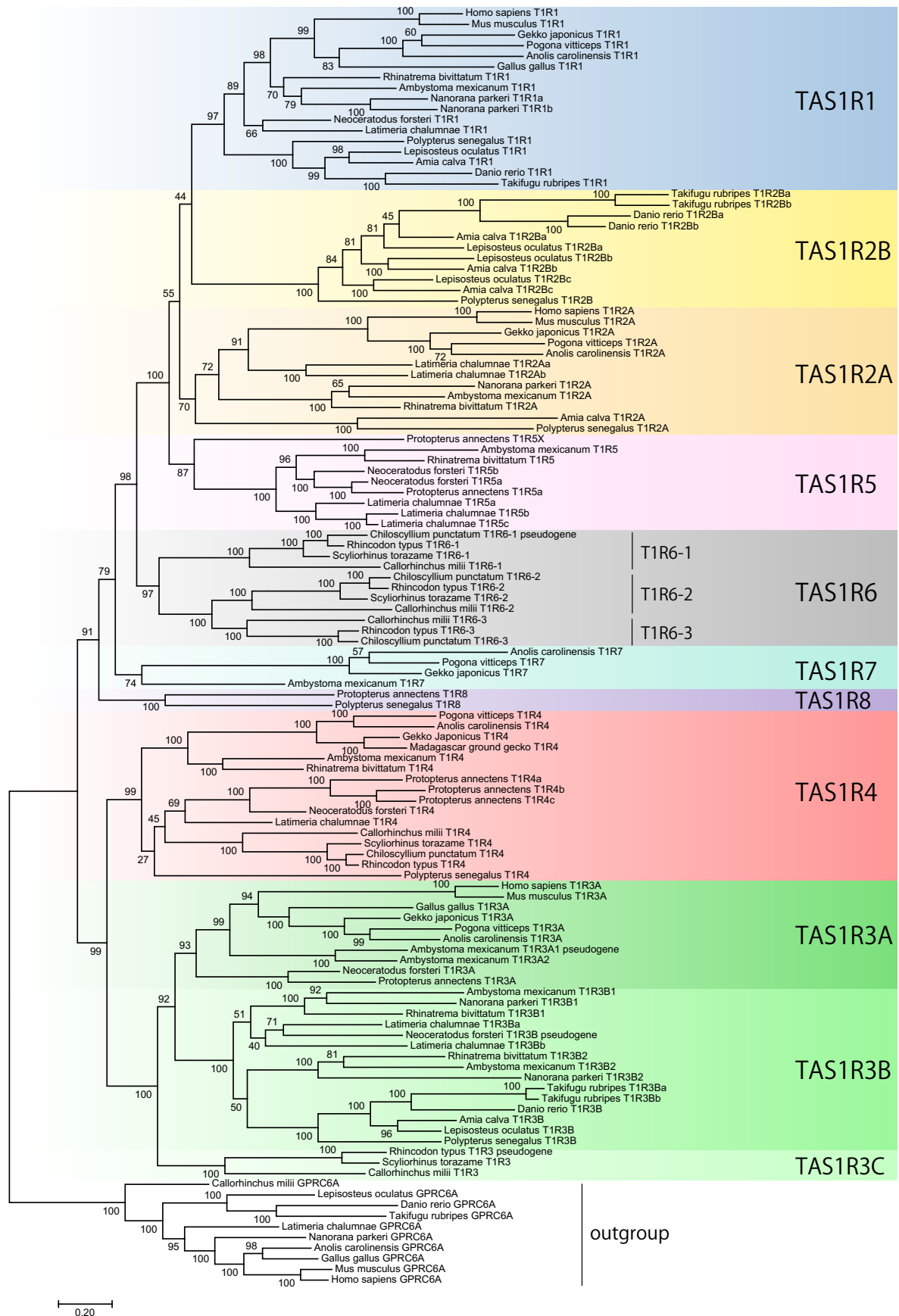
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third

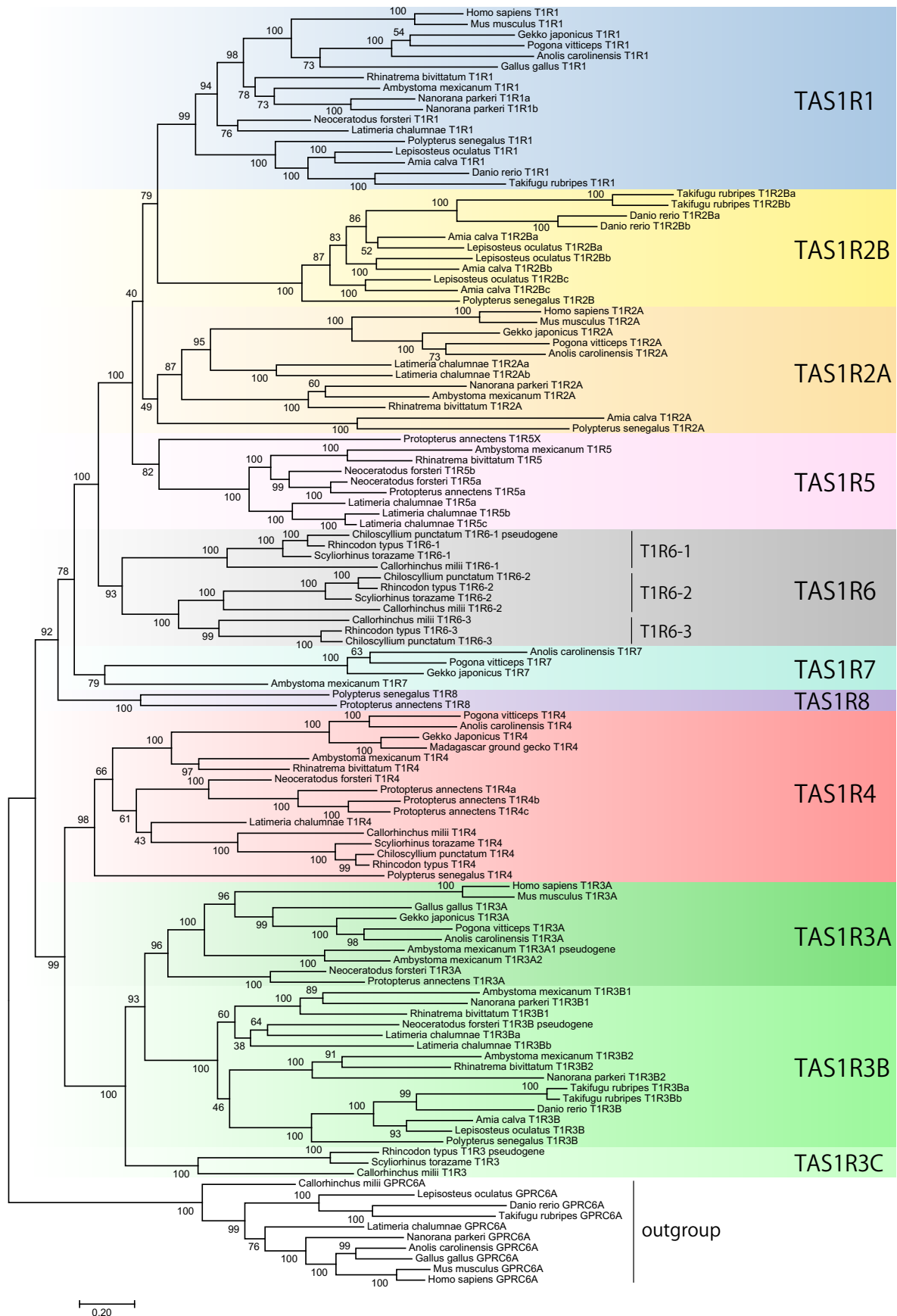
party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023



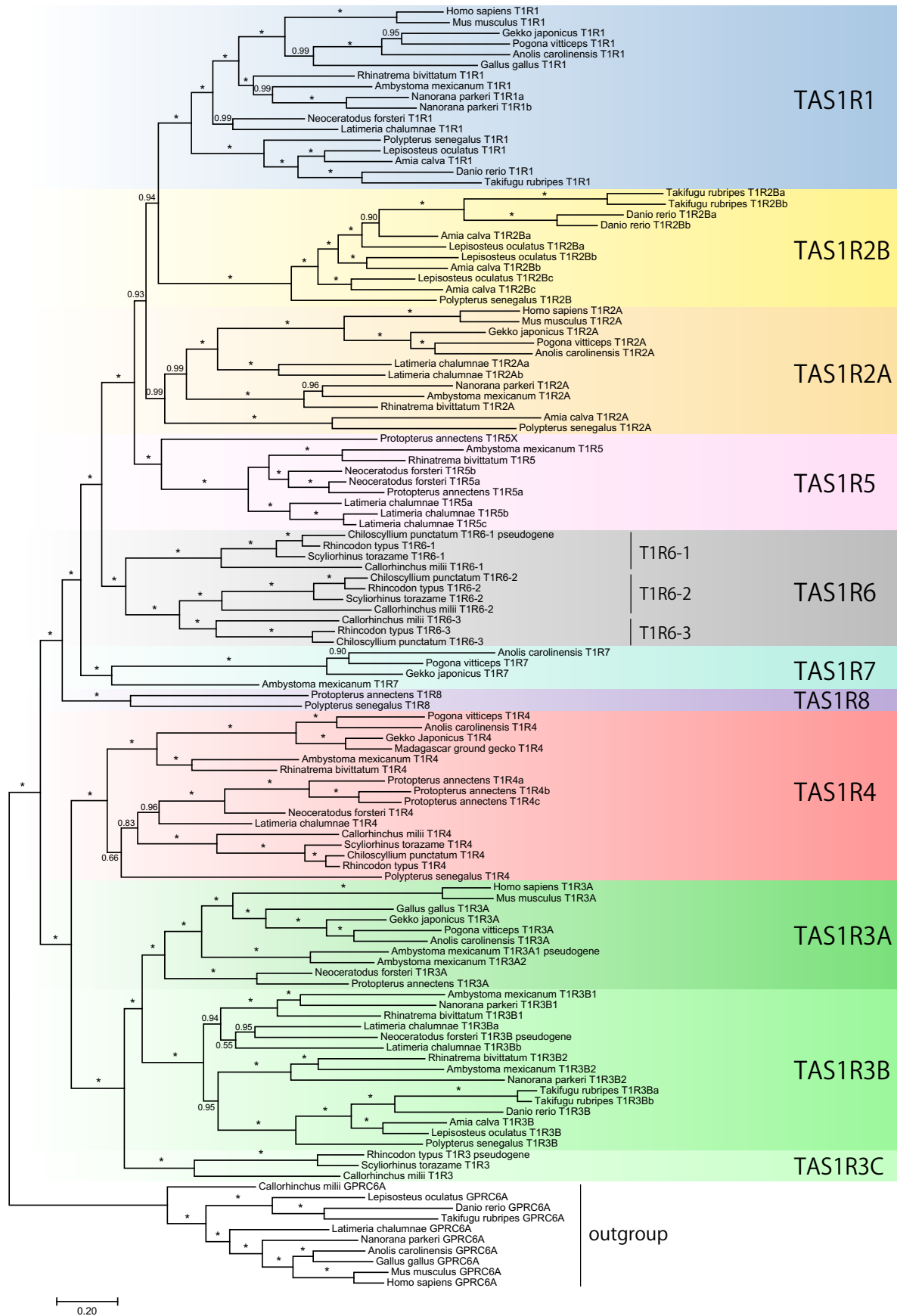
Extended Data Fig. 1 | Maximum-likelihood tree of *TASIR* members identified for 21 vertebrates. A maximum-likelihood tree was constructed from the amino acid sequences encoded by *TASIRs* using RAXML with the

JTT + G (CAT approximation) model. Branch supports represent bootstrap values calculated with 1,000 replications. *TASIR* clade names are shown on the right.



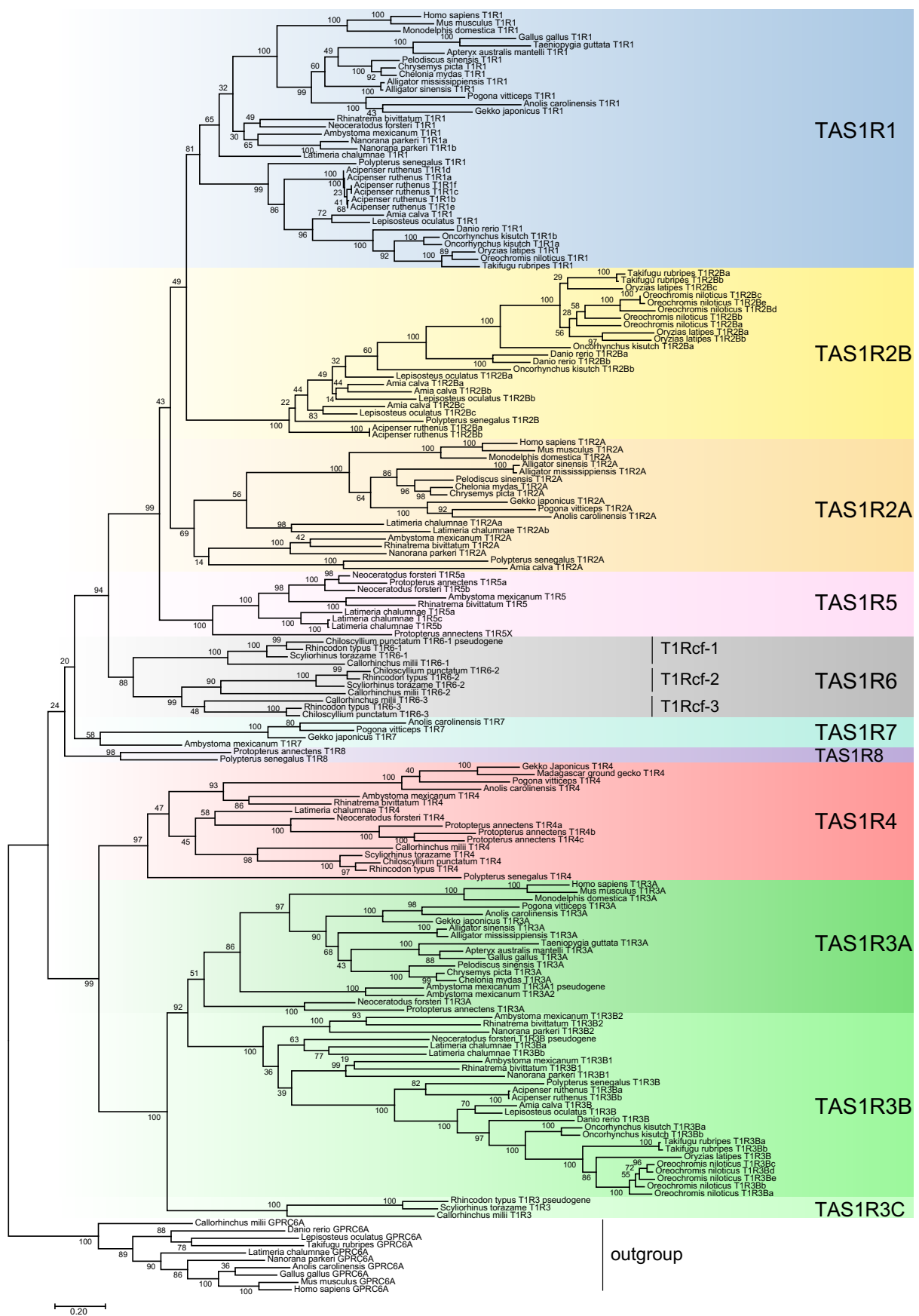
Extended Data Fig. 2 | Maximum-likelihood tree with the mixture model for the *TASIR* members identified for 21 vertebrates. A maximum-likelihood tree was constructed from the amino acid sequences encoded by *TASIRs* using IQ-tree

under the posterior mean site frequency approximation of the JTT + C20 + F + I model. Branch supports represent bootstrap values calculated with 1,000 replications. *TASIR* clade names are shown on the right.



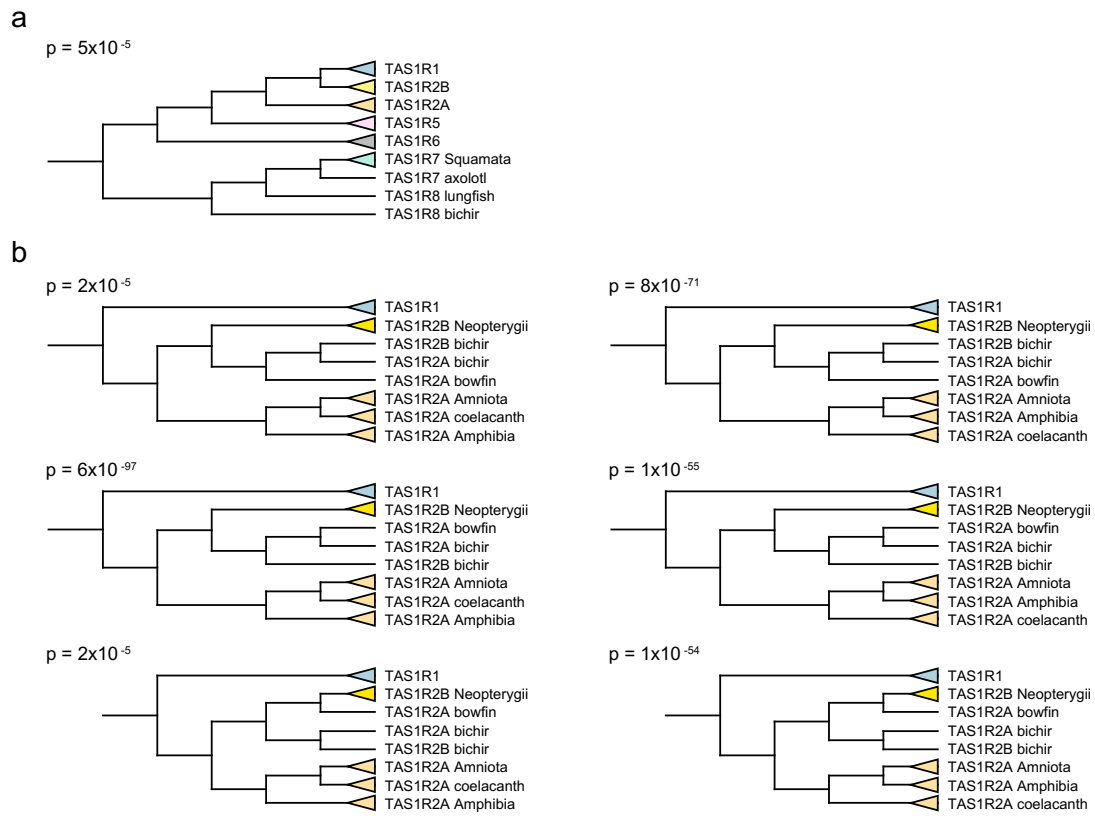
Extended Data Fig. 3 | Bayesian tree of *TAS1R* members identified for 21 vertebrates. Bayesian tree inference was performed for the amino acid sequences encoded by *TAS1Rs* using MrBayes with the JTT-F + Γ_4 model.

Branch supports represent Bayesian posterior probabilities, and asterisks indicate a posterior probability of 1.00. *TAS1R* clade names are shown on the right.



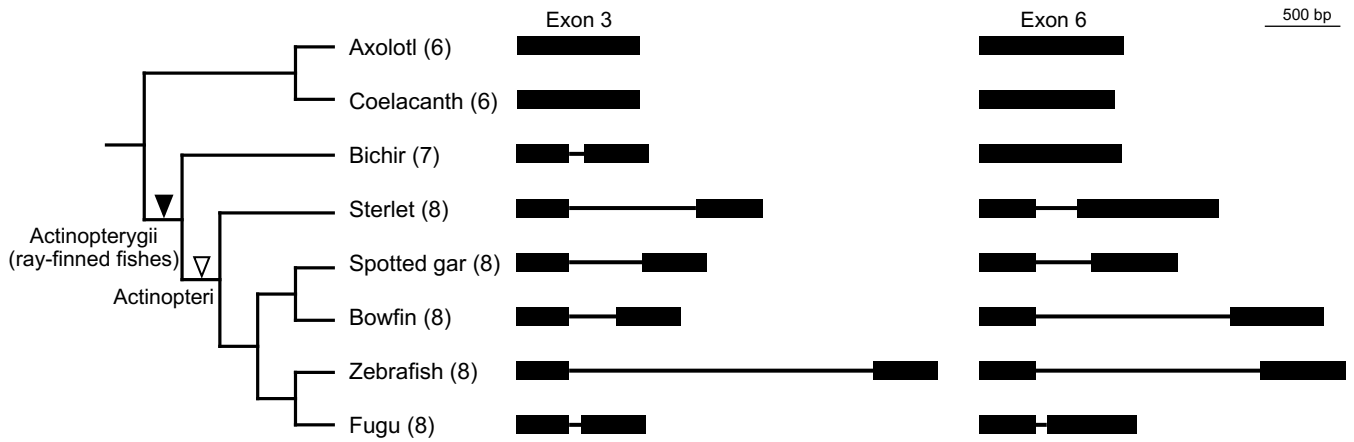
Extended Data Fig. 4 | Maximum-likelihood tree for exon 6 sequences of TASIRs identified for 33 vertebrates. A maximum-likelihood tree was constructed from the amino acid sequences of TASIR exon 6 using RAxML with

the JTT + G (CAT approximation) model. Branch supports represent bootstrap values calculated with 1,000 replications. TASIR clade names are shown on the right.



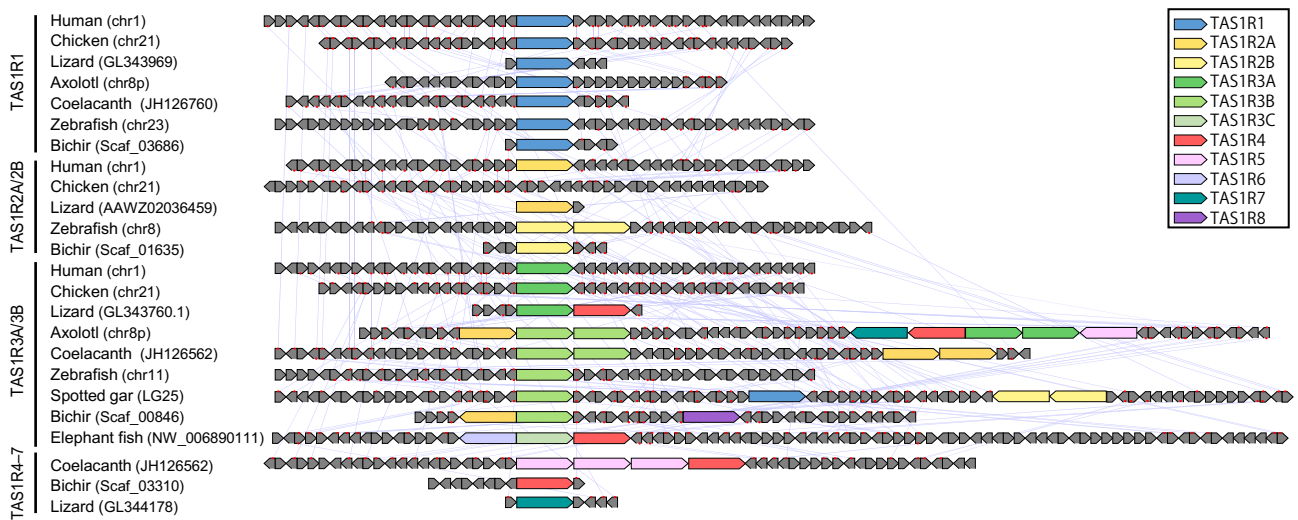
Extended Data Fig. 5 | Tree topologies examined for the approximately unbiased test. a, Tree topology assuming the grouping of *TAS1R7* and *TAS1R8* and showing a species tree. **b**, Tree topologies for various relationships among *TAS1R2A* and *TAS1R2B* genes from amphibians, coelacanth, and ray-finned fishes.

The p -values for the approximately unbiased test calculated with CONSEL are shown above each tree. Relationships within each collapsed group were fixed to be the same as in the maximum-likelihood tree (Fig. 1).



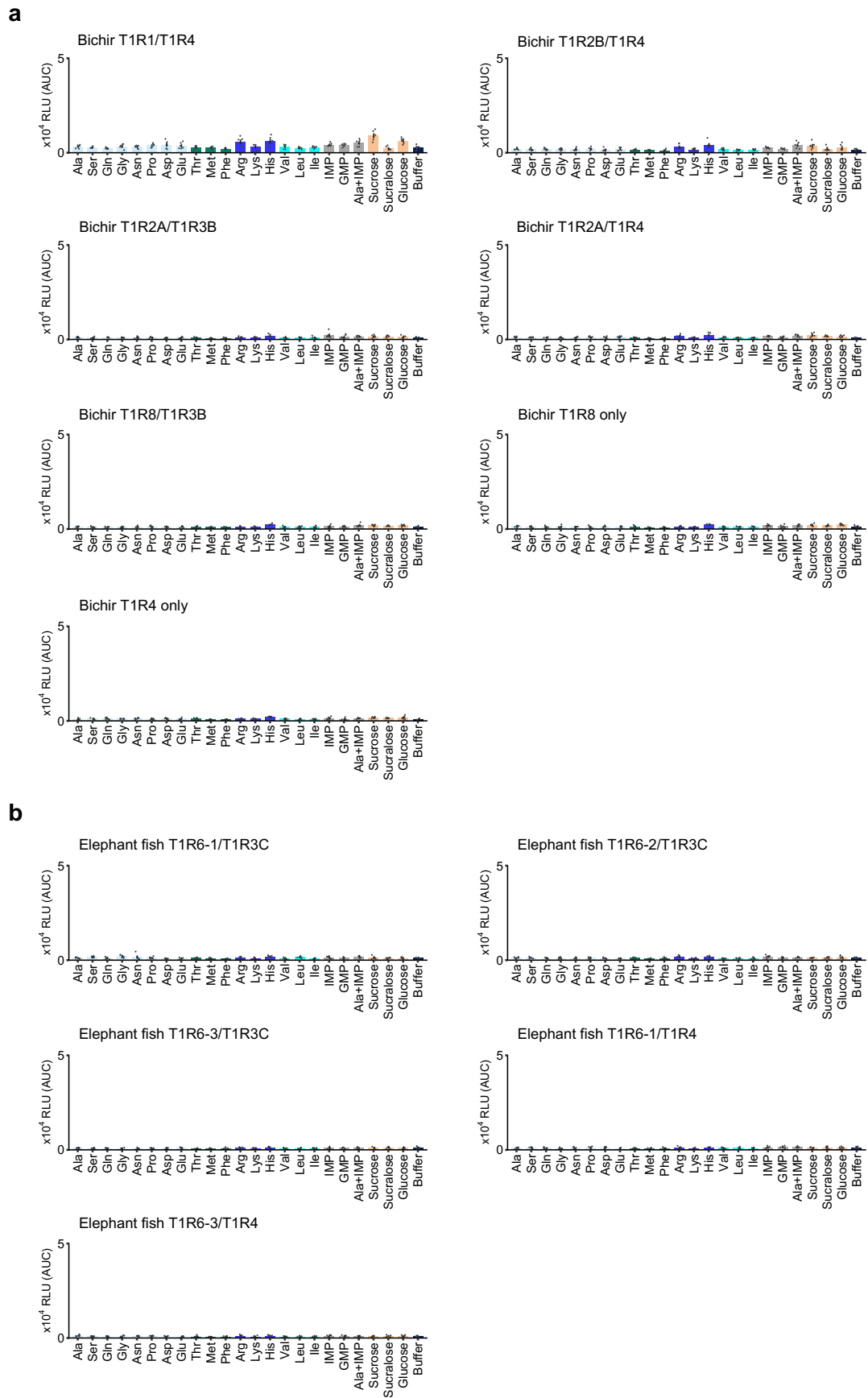
Extended Data Fig. 6 | The history of intron acquisitions for the two exons of *TASIR3B* in ray-finned fishes. Filled boxes represent exons, and lines represent introns. The phylogenetic relationship among *TASIR3B*-containing species is shown on the left. The closed and open arrowheads indicates the deduced

timing (*that is*, the common ancestors of Actinopterygii and Actinopteri) for the acquisitions of introns in exons 3 and 6 of *TASIR3B*, respectively. The total number of exons of the *TASIR3B* genes is shown in parentheses.



Extended Data Fig. 7 | Comparison of synteny among vertebrates. *TASIRs* and non-*TASIR* genes are represented by colored and grey polygons, respectively, with each pointed end indicating the direction of transcription. Genomic regions

are categorized according to the *TASIR1*-, *TASIR2*-, *TASIR3*-containing regions as well as the other *TASIR*-containing regions. Orthologous non-*TASIR* genes between the closely represented species are connected by light-blue lines.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | No significant response of various combinations of TIRs. **a**, Five TIR combinations, TIR4-only, and TIR8-only from the bichir were coexpressed in HEK293T cells, and their responses to each of the 17 amino acids (50 mM), nucleic acids (10 mM), sugars and sucralose (100 mM) were tested. Values represent the mean \pm standard error of six independent experiments

performed with duplicate samples. **b**, Same as **(a)** except using five TIR combinations from the elephant fish. A lack of response may be due to technical issues with the heterologous expression system, or there is a possibility that they were functional and could respond to ligands not used in this study.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a | Confirmed |
|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

- | | |
|-----------------|---|
| Data collection | All publicly-available genome information from the National Center for Biotechnology Information and Ensembl, which was used for data collection, has been described and referenced in the manuscript. |
| Data analysis | blastn, tblastn, Bridger ver. r2014-12-01, AUGUSTUS ver. 3.2.3, bowtie2, PREQUAL, MAFFT ver. 7.427, RAxML ver. 8.2.12, IQ-TREE ver. 2.2.2.6, MrBayes 3.2.6, iTOL, CONSEL v0.20, TRANSFAC v11.3, STAMP, MEME |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The sequences and phylogenetic trees have been provided in Supplementary Data.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Use the terms *sex* (biological attribute) and *gender* (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.

Reporting on race, ethnicity, or other socially relevant groupings

Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status). Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.) Please provide details about how you controlled for confounding variables in your analyses.

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

The functional results were obtained from an appropriate number of independent experiments (six) of independent experiments performed with duplicate samples.

Data exclusions

No data were excluded from the analyses.

Replication

All attempts to reproduce the experiments were successful.

Randomization

Not relevant because no data analysis requiring randomization was performed.

Blinding

Not relevant because no data analysis requiring blinding was performed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involvement
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	Polypterus senegalus
Wild animals	This study did not involve wild animals.
Reporting on sex	The study utilized both male and female bichirs and observed no differences between the sexes.
Field-collected samples	This study did not include samples collected from the field.
Ethics oversight	This study was carried out in accordance with the National Institutes of Health guide for the care and use of Laboratory animals (NIH Publications No. 8023, revised 1978).

Note that full information on the approval of the study protocol must also be provided in the manuscript.