Diet Shapes the Evolution of the Vertebrate Bitter Taste Receptor Gene Repertoire

Diyan Li\textsuperscript{1,2} and Jianzhi Zhang\textsuperscript{*2}
\textsuperscript{1}College of Animal Science and Technology, Sichuan Agricultural University, Ya’an, Sichuan, China
\textsuperscript{2}Department of Ecology and Evolutionary Biology, University of Michigan
\textsuperscript{*}Corresponding author: E-mail: jianzhi@umich.edu.

Abstract

Vertebrate Tas2r taste receptors bind to bitter compounds, which are typically poisonous, to elicit bitter sensation to prevent the ingestion of toxins. Previous studies noted a marked variation in the number of Tas2r genes among species, but the underlying cause is unclear. To address this question, we compile the Tas2r gene repertoires from 41 mammals, 4 birds, 2 reptiles, 1 amphibian, and 6 fishes. The number of intact Tas2r genes varies from 0 in the bottlenose dolphin to 51 in the Western clawed frog, with numerous expansions and contractions of the gene family throughout vertebrates, especially among tetrapods. The Tas2r gene number in a species correlates with the fraction of plants in its diet. Because plant tissues contain more toxic compounds than animal tissues do, our observation supports the hypothesis that dietary toxins are a major selective force shaping the diversity of the Tas2r repertoire.

Key words: bitter taste receptor, Tas2r, herbivore, carnivore, omnivore.

Mammals can detect five basic tastes: sweet, salty, sour, bitter, and umami (Kinnaman and Cummings 1992; Lindemann 1996). Among them, the bitter taste is thought to help prevent the ingestion of poisonous substances such as plant alkaloids, because poisons typically taste bitter (Garcia and Hankins 1975; Glendinning 1994). The bitter sensation is mediated by a group of seven-transmembrane-domain G-protein-coupled receptors known as Tas2rs, which are encoded by members of the Tas2r gene family (Adler et al. 2000; Chandrashekar et al. 2000; Matsunami et al. 2000). Each Tas2r is responsive to several bitter compounds, whereas different Tas2rs show different sensitivities to the same bitter compounds (Meyerhof et al. 2010). The Tas2r repertoire, described thus far in 17 vertebrates on the basis of genome sequences (table S1, Supplementary Material online), varies greatly in size among species (Conte et al. 2002 2003; Shi et al. 2003; Fredriksson and Schiöth 2005; Go 2006; Lagerstrom et al. 2006; Shi and Zhang 2006; Gloriam et al. 2007; Dong et al. 2009; Shi and Zhang 2009; Jiang et al. 2012). Frequent Tas2r gains and losses in evolution were also noted in many analyses of individual loci (Parry et al. 2004; Wang et al. 2004; Fischer et al. 2005; Go et al. 2005; Sugawara et al. 2011). The underlying reason of this variation, however, is unclear. Because toxins are more abundant in plant tissues than in animal tissues (Glendinning 1994; Wang et al. 2004), herbivores should face a stronger selective pressure than carnivores to detect poisonous food. Given that different Tas2rs can detect different bitter compounds (Meyerhof et al. 2010), it is reasonable to assume that gains of Tas2rs via gene duplication would generally increase the number of detectable toxins, whereas Tas2r losses would reduce this number. Thus, we predict more functional Tas2r genes in herbivores than in carnivores. Here we test this hypothesis after identifying Tas2r genes from 54 vertebrates.

We used previously described full-length Tas2rs from the human, mouse, chicken, and zebrafish as queries to identify Tas2r genes from the genome sequences of 54 vertebrates, including the 17 species previously analyzed (see Materials and Methods). Because Tas2r genes lack introns in coding regions and have on average ~300 codons, gene identification was straightforward. The 54 species include 41 mammals, four birds (chicken, turkey, zebra finch, and medium ground finch), two reptiles (a turtle and a lizard), one amphibian (western clawed frog), and six fishes (five teleosts and a coelacanth; fig. 1). We divided the identified Tas2r genes into three categories. Intact genes refer to those with at least 270 amino acids, start codon, stop codon, and seven transmembrane domains. Partial genes refer to those that have at least 100 codons and have either a start or a stop codon but not both; their open reading frames are truncated because of incomplete genome sequencing. Pseudogenes refer to those that have at least 300 nucleotides, but the open reading frame is interrupted by premature stop codons and/or frame-shifting mutations. The total number of Tas2r genes of all three categories varies substantially among species, with the largest number (69) found in the guinea pig and the smallest (3) in the chicken, turkey, and stickleback (fig. 1). When only intact Tas2r genes are concerned, the largest number (51) is in the frog, while the smallest (0) is in the dolphin (fig. 1). The proportion of pseudogenes in the Tas2r repertoire ranges from 0% in the chicken and the five teleosts to 100% in the dolphin (fig. 1).

We aligned the amino acid sequences of all 856 intact Tas2r genes from 53 species (dolphin has no intact Tas2r). We constructed a neighbor-joining tree of these genes (fig. 2;
Fig. 1. The *Tas2r* gene repertoires of 54 vertebrates determined in this study. See Materials and Methods for the sources of the species tree and divergence times. Dietary information is from various sources (table S2, Supplementary Material online). C, carnivorous; H, herbivorous; O, omnivorous.
Fig. 2. Evolutionary relationships of all 856 intact Tas2r genes from 53 vertebrates (dolphin has no intact Tas2r). The tree is reconstructed using the neighbor-joining method with protein Poisson-corrected gamma distances and is rooted with a fish V1r gene (GenBank: AB670529.1). Branch lengths are drawn to scale, which is measured by the number of amino acid substitutions per site. See figure S1 (Supplementary Material online) for the detailed tree with species and gene names and bootstrap percentages. The blue asterisk indicates the lineage from which all mammalian, avian, and turtle Tas2rs are derived.
Diet impacts the size of the vertebrate Tas2r repertoire (fig. S1, Supplementary Material online), using a fish V1r gene as an outgroup, because V1r genes are known to be the closest relative to Tas2r genes (Shi et al. 2003). The bootstrap values in the tree are generally low (fig. S1, Supplementary Material online) because of the relatively small number of aligned gap-free sites. We did not include the partial Tas2r genes or pseudogenes in the phylogenetic analysis, because their inclusion would drastically reduce the already low number of gap-free sites. The obtained gene tree suggests a major division between the Tas2r genes of fishes (light blue lineages in fig. 2) and tetrapods (all other colors in fig. 2, Supplementary Material online). Further, several basal lineages of Tas2r include only genes from the fishes, frog, and lizard, whereas all mammalian, avian, and turtle Tas2r genes appear to have originated from only one basal lineage (marked with an asterisk in fig. 2). It was noted a decade ago in a comparison of human and mouse Tas2rs that some lineages of Tas2rs are enriched with species-specific gene duplications, whereas other lineages are relatively duplication free (Shi et al. 2003). This dichotomy is also evident in the present tree, as some lineages show a cluster of genes from the same species or group of closely related species (marked with one color), whereas other lineages show genes from distantly related species (marked with many colors; fig. 2).

To investigate the gains and losses of Tas2rs in vertebrate evolution, we estimated the numbers of intact Tas2r genes in ancestral species and mapped gene gains and losses onto the species tree, using the reconciled-tree method (Page and Charleston 1997). Because the method is computationally intensive, we chose 32 species to represent all major evolutionary lineages covered by the full set of 54 species. Based on this inference (fig. S2, Supplementary Material online), the intact Tas2r gene repertoire was relatively small (<10 genes) in the common ancestor of vertebrates, that of tetrapods, and that of mammals. Only in the common ancestor of therians did the intact gene number exceed 10. Gains and losses of Tas2r genes were fairly common throughout vertebrate evolution. In particular, massive (>10) gene gains occurred in the lineages leading to the frog, lizard, microbat, rabbit, guinea pig, and the common ancestor of mouse and rat. Massive (>10) gene losses were observed in the lineage leading to the rabbit and that to manatee. Dolphin is not included in this analysis due to its lack of any intact Tas2r, but massive gene losses apparently occurred in dolphin (Jiang et al. 2012), because it has only Tas2r pseudogenes and because cow, its closest relative in our data set, has 21 intact genes. The lost genes in these massive losses appear to be randomly distributed among sublineages of the Tas2r family.

To examine the potential impact of diet on Tas2r repertoire evolution, we categorized vertebrates into carnivores, omnivores, and herbivores (fig. 1), on the basis of numerous references (table S2, Supplementary Material online), which often cited the 90% rule (Harestad and Bunnell 1979). That is, a species is considered herbivorous (or carnivorous) if its diet comprises 90% or more plant (or animal) tissues; all other vertebrates are considered omnivorous. We coded the dietary preference of a species by 0 (carnivorous), 0.5 (omnivorous), or 1 (herbivorous), and then correlated the dietary code of a species with properties of its Tas2r repertoire. Because of the phylogenetic nonindependence among the vertebrates analyzed, we employed phylogenetically independent contrasts (PICs; Felsenstein 1985b) in our regression analysis. That is, we converted the 54 phylogenetically correlated data points into 53 PICs, using the information of the species tree of the 54 species including their divergence times (see Materials and Methods). Supporting our hypothesis that consuming plants (rather than animals) demands more Tas2r genes, the PICs in the dietary code and that in the Tas2r gene repertoire size (i.e., the total number of intact genes, partial genes, and pseudogenes) are positively correlated ($R = 0.429$, $P < 0.001$ in one-tail $t$-test; fig. 3A). The same is true when only intact and partial genes are considered ($R = 0.265$, $P = 0.027$ in

![Fig. 3.](image-url) Diet impacts the size of the vertebrate Tas2r repertoire. (A) Phylogenetically independent contrast (PIC) in total number of intact genes, partial genes, and pseudogenes of the Tas2r repertoire is significantly positively correlated with that in diet code. (B) PIC in total number of intact and partial Tas2r genes is significantly positively correlated with that in diet code. The diet code is 0, 0.5, and 1 for carnivores, omnivores, and herbivores, respectively.
Bivores cannot “afford” rejecting all bitter foods due to the evolution are complex. Evolutionary patterns of taste receptor genes that are consistent with the expectations from the ecology of the organisms have been reported (Wang et al. 2004; Zhao, Yang, et al. 2010; Zhao, Zhou, et al. 2010; Jiang et al. 2012), but inconsistent patterns also abound (Zhao et al. 2012; Zhao and Zhang 2012). It is important to mention that some taste receptor genes play unexpected roles in addition to their canonical functions. For example, mouse Tas2rs are also used by nasal chemosensory cells to detect irritants and bacterial signals (Tizzano et al. 2010). Mouse Tas1r3, responsible for sweet and umami tastes, is also found in the testis, and the deletion of Tas1r3 causes male sterility (Mosinger et al. 2013). Given these complications, to avoid spurious results, it is imperative to examine a diverse group of species when testing the potential impact of an ecological factor on taste receptor gene evolution. Further, because of the existence of multiple factors, the impact of any factor is likely to be quantitative rather than qualitative, and a small number of counterexamples should not automatically refute the potential impact of a factor in general.

Materials and Methods

We used 25 human, 34 mouse, 3 chicken, and 4 zebrafish Tas2rs retrieved from GenBank as queries to conduct TBlastN (Altschul et al. 1990) searches (e-value cutoff = 1e−10) in each of the vertebrate genome available at the University of California–San Cruz (UCSC) genome browser (http://genome.ucsc.edu/index.html, last accessed November 20, 2013) and Ensembl (http://www.ensembl.org/index.html, last accessed November 20, 2013) in October 2012. We followed a previous study (Shi and Zhang 2006) in identifying Tas2r genes. Briefly, candidate Tas2r genes identified via TBlastN were verified by the TransMembrane prediction using Hidden Markov Models (TMM/MM) method for the presence of seven transmembrane domains (Krogh et al. 2001) and were examined by BlastP searches against the entire GenBank to ensure that the best hit with an annotation is a known Tas2r gene. For the sheep, elephant, Nile tilapia, medaka, and budgerigar, the identified Tas2r sequences contain numerous ambiguous nucleotides due to low sequencing quality. After excluding these species, we analyzed the Tas2r repertoires from 54 genomes (fig. 1). The protein sequences of all intact Tas2rs are provided in supplementary data set S1, Supplementary Material online. Although the genome sequence coverage varies among the 54 species, the coverage is not expected to differ according to the diet of the species. Thus, our analysis of the dietary impact on Tas2r repertoire evolution is not expected to be affected by different coverages or genome assemblies.

The deduced Tas2r sequences were aligned using Clustal X (Chenna et al. 2003) with manual adjustments. A neighboring tree (Saitou and Nei 1987) of 856 protein sequences of intact Tas2rs was constructed using MEGAS (Tamura et al. 2011) with Poisson-corrected gamma distances (shape parameter = 1; fig. 2; fig. S1, Supplementary Material online). The reliability of the estimated tree was evaluated by the bootstrap method (Felsenstein 1985a) with 1,000 replications. Percentage bootstrap values ≥50 are shown above branches (fig. S1, Supplementary Material online).

We used the package Analyses of Phylogenetics and Evolution (APE; Paradis et al. 2004) to conduct a PIC analysis (Felsenstein 1985b). The tree shown in figure 1 was used. The topology of the tree was downloaded from the UCSC Genome Browser (http://hgdownload-test.cse.ucsc.edu/goldenPath/mm10/multiz60way/mm10.60way.commonNam es.nh, last accessed November 20, 2013), whereas the branch
lengths were based on multiple sources (table S3, Supplementary Material online).

We used the reconciled-tree method to infer gains and losses of Tas2rs in the vertebrate phylogeny (Gorecki et al. 2011). Nodes with <50 bootstrap percentages in the estimated gene tree were collapsed before this inference. The species tree used was the same as shown in figure 1. We also used the reconciled-tree method without collapsing any nodes in the gene tree and found the estimated gene numbers for ancient nodes (fishes, frog, turtle, birds, platypus, and opossum) of the reconciled tree to be largely unaltered, but those for recent nodes tended to vary. Hence, collapsing weakly supported nodes, as was done for figure S2 (Supplementary Material online), provides conservative estimates of the numbers of gene gains and losses.

Supplementary Material
Supplementary data set S1, figures S1 and S2 and tables S1–S3 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments
The authors thank Xiaoshu Chen, David Glenn Smith, Jian-Rong Yang, and four anonymous reviewers for valuable comments. This work was supported in part by a research grant from the Chinese Agriculture Research System (CARS-41) to Qing Zhu and a research grant (GM080285) from the U.S. National Institutes of Health to J.Z.

References


