PNAS

Supporting Information for

Evolution of bitter receptor genes and ontogenetic dietary shift in a frog

Xiangyu Hao, Hengwu Jiao, Dahu Zou, Qiyang Li, Xiangqun Yuan, Wenbo Liao, Peihua Jiang, Huabin Zhao

Correspondence to: Huabin Zhao
Email: huabinzhao@whu.edu.cn

This PDF file includes:

Supporting text: Extended Methods
Legends for Datasets S1 to S6
SI References

Other supporting materials for this manuscript include the following:

Datasets S1 to S6

Supporting Information Text

Extended methods

Identification of *Tas2r* genes of Anura. In order to search for *Tas2r*s in the amphibian order Anura, we obtained all 23 reference genomes of Anura species from GenBank, and discarded three low-quality genomes with Scaffold N50 shorter than 2000-bp (Dataset S1). Previously determined *Tas2r* intact protein sequences of vertebrates (1, 2) were used as queries, and the tblastn program was used to align and search against the anuran genomes with an E-value cutoff of $1e^{-}10$. The tblastn hits with a length longer than 270 amino acids were retained for subsequent analysis following previous studies (1, 2), and both ends of each hit were extended by 500-bp to determine the correct start and stop codons. We identified intact genes that were characterized by an intact open reading frame and a full-length coding region. Next, the TMHMM method (3) was used to predict transmembrane domains of the identified intact genes to ensure that they have the complete seven transmembrane domains to perform normal bitter taste functions. Finally, we used the blastp program to check whether the best hit is a bitter receptor gene, and determined the final set of bitter receptor genes. All *Tas2r* gene sequences of Anura species are provided in Dataset S2. For convenience, we named each *Tas2r* gene numerically following the order in which they were identified.

Moreover, we reconstructed a *Tas2r* gene tree of vertebrates to ensure phylogenetic affinities of the newly identified genes. The data set included published genes obtained from Li and Zhang (2014) (2) and newly identified genes from 20 frog species in this study. Multiple sequence alignment of all deduced protein sequences was performed using the L-INS-I strategy in MAFFT v7.123b (4, 5), and poorly aligned regions were removed using Gblocks v0.91 (6). The fish vomeronasal pheromone receptor gene *V1R3* (GenBank accession number: BAM35766) was selected as the outgroup. ModelFinder (7) was used to predict the optimal nucleotide substitution model. Phylogenetic reconstruction of *Tas2r*s was carried out using the maximum likelihood (ML) method in IQ-TREE v2.1.4 (8). Bootstrap values of each node were evaluated via the Ultrafast Bootstrap approximation algorithm with 10,000 replicates (9). Finally, the phylogenetic tree was visualized using the online website iTOL (10). The phylogenetic tree showed that six species of fish, as the basal clade of the whole tree, are the sister groups to frogs and other vertebrates (Dataset S5), suggesting that our newly identified genes from 20 frog genomes are indeed *Tas2r* genes.

Sample collection, RNA isolation, library preparation and sequencing. Six similarly sized tadpoles and six similarly sized adult American bullfrogs were sampled in a farm from Nanchong City, Sichuan Province, China, where this species was raised for food, in August 2020. All individuals were euthanized in the lab, and their taste tissues were subsequently sampled. The selected taste tissue of an adult bullfrog contained the tongue's upper surface and oral epithelium, whereas that of tadpoles included only oral epithelium, as a tongue was not observed at the tadpole life stage. Six individuals of tadpoles and six individuals of adult bullfrogs were used as biological replicates for transcriptome analysis. Total RNAs were isolated from the taste tissues using TRIzol™ Reagent (Invitrogen Corporation, CA, USA) following the manufacturer's protocol. The quality and quantity of RNA samples were assessed using a NanoDrop (NanoDrop Technologies, DE, USA) instrument. Total RNAs were sent to Origingene Biomedical Technology Co. Ltd. (Shanghai, China) for RNA-seq library preparation using the Illumina Truseq RNA Sample Preparation Kit (San Diego, CA, USA). Details of library preparation are previously described (11). Twelve libraries were sequenced to produce 125-bp paired-end reads on the Illumina HiSeq 2500 platform. All experimental protocols on bullfrogs were reviewed and approved by the Ethics Committee of Wuhan University.

Transcriptome assembly and differential gene expression analysis. The transcriptome analysis workflow followed a previous study (12). The raw data of paired-end sequencing was quality-controlled and filtered using Fastp v0.23.0 (13), and reads with a Phred quality score less

than 40 (q < 40) were removed. Clean data was mapped to the American bullfrog reference genome (GenBank assembly accession: GCA_002284835.2) using Hisat2 v2.1.0 (14, 15) with default parameters. SAMtools v1.4 (16) was used to sort and convert the alignments generated by Hisat2 into BAM format files. Transcript assembly and quantification were performed using StringTie v2.1.4 (17). Differential gene expression analysis was performed in the R package DESeq2 (18), using a generalized linear model with the Wald statistical test by assuming that underlying gene expression data were distributed following a negative binomial distribution. Differentially expressed genes (DEGs) were identified between tadpoles and adults. To minimize sequencing noise, our differential expression analysis used the threshold as follows: absolute fold change greater than 2, and adjusted p-value smaller than 0.05. This method could substantially reduce the impact of the variability in the read counts among individual samples (18), although such variability seems rather large (Fig. 2A). Eventually, a total of 33 differentially expressed *Tas2r*s in the American bullfrog were identified, using the above threshold, of which 11 were preferentially expressed in tadpoles and 22 were preferentially expressed in adults. Specifically, the 11 preferentially expressed genes in tadpoles refer to those showing a significantly higher expression level in tadpoles than in adults, whereas the 22 preferentially expressed genes in adults showed a significantly higher expression level in adults than in tadpoles. Heatmaps of *Tas2r* expression levels were constructed using TBtools based on normalized transcripts per million (TPM) values using the zero-to-one method at row scale (19). The raw TPM values are also shown in Dataset S3.

Selection criteria of functionally tested *Tas2r* genes. In order to select representative genes for functional experiments, we constructed a ML phylogenetic tree based on the Ultrafast Bootstrap approximation algorithm with 10,000 replicates (Dataset S4), and then selected the DEGs based on their phylogenetic positions, because genes with different clustering relationships may represent different functions. Eventually, a total of 11 DEGs were selected, including 6 preferentially expressed in tadpoles (*Tas2r1*, *Tas2r26*, *Tas2r32*, *Tas2r55*, *Tas2r99* and *Tas2r127*) and 5 preferentially expressed in adults (*Tas2r16*, *Tas2r38*, *Tas2r137*, *Tas2r143* and *Tas2r175*) (Dataset S4).

Bitter compounds. Most bitter compounds in nature come from plants, and the main categories contained in plants are glycosides or alkaloids (20, 21). The bitter compound collection in this study contains 10 representatives of natural compounds from plants (amygdalin, arbutin, camphor, chloramphenicol, papaverine hydrochloride, picrotoxinin, quinine hydrochloride dihydrate, D-salicin, helicin and yohimbine hydrochloride) (Dataset S6). These substances have been determined to be bitter and were recorded in BitterDB (22). Unfortunately, we cannot completely differentiate bitter substances derived from the diets between tadpoles and adult bullfrogs. In addition, nearly no literature provided evidence on bitter receptors that could sense bitter compounds uniquely derived from insects. Based on these limitations, this study only selected bitter compounds from plants to explore the dietary differences between tadpoles and adults. However, it is worth noting that insects often sequester secondary metabolites (bitter compounds) from consumed plants to serve as their own defense systems (23, 24), thus insectivorous adult bullfrogs must encounter many bitter compounds in their insect diets initially derived from plants. All compounds in this study were purchased from Sigma-Aldrich. The compounds were dissolved in Dulbecco's phosphate-buffered saline (DPBS) or a mixture of dimethyl sulfoxide (DMSO) at a final DMSO concentration of <0.1% to protect transfected cells from toxicity. The highest concentrations of bitter compounds used in our study were based on Meyerhof et al. (2010) (25).

Construction of Tas2r expression plasmids. The full-length coding sequences of *Tas2r*s were sent to GenScript (Nanjing, China) for chemical synthesis, afterwards they were inserted into the expression vector pEAK10 with the Kozak sequence introduced before the start codon. All constructs were verified by Sanger sequencing.

Functional assays of Tas2rs. Human embryonic kidney 293 (HEK293) cells (PEAKrapid) were purchased from ATCC (American Type Culture Collection) and cultured in Opti-MEM supplemented

with 5% fetal bovine serum (FBS). The cells were seeded uniformly in 96-well plates at a density of 45,000–50,000 per well. After 24–26 hours, the HEK293 cells were transiently transfected by a Tas2r construct (0.1 μg/well) with a coupling chimeric G protein Gα16-gust44 (0.1 μg/well) using Lipofectamine 2000 (0.5 μL/well). Cells transfected with Gα16-gust44 only were used as negative controls. After six hours of expression, the serum-free medium was changed to Opti-MEM with 5% FBS. After 24 hours of transfection, the culture medium was removed. Cells were washed once with DPBS buffer, and then loaded with the calcium-sensitive dye Fluo-4 AM (2.5 μM, Invitrogen). One hour later, the dye solution Fluo-4 AM was removed and cells were washed two more times with DPBS buffer and placed in the dark for 30 minutes. Finally, cells were assayed in a FlexStation III reader (Molecular Devices) to monitor the fluorescence changes every two seconds for two minutes. After 30 minutes of reading, we added a bitter compound solution to record the change of relative fluorescence units (excitation at 488 nm, emission at 525 nm, cut-off at 515 nm). All measurements were independently repeated at least three times. Calcium mobilization was quantified as the ratio of the changes in fluorescence (ΔF) relative to F (the peak of fluorescence minus baseline fluorescence) from triplicate experiments. Calcium signal traces and bar graphs were generated using GraphPad Prism 9, and Student's *t*-tests were used for statistical analysis (*P< 0.05, **P< 0.01, ***P< 0.001). In these assays, human Tas2r16 was selected as the positive control, because it has been determined to be sensitive to many known bitter ligands (26, 27).

Legend for Dataset S1 (separate file). Number of intact *Tas2r* genes in vertebrates.

Legend for Dataset S2 (separate file). All newly determined *Tas2r* nucleotide sequences of Anura (frogs and toads) in this study.

Legend for Dataset S3 (separate file). TPM (transcripts per million) values of the American bullfrog *Tas2r* genes in differential gene expression analysis.

Legend for Dataset S4 (separate file). Phylogenetic tree of the 180 *Tas2r* genes identified from the genome sequence of the American bullfrog, rooted using a fish *V1R3* (accession number: BAM35766).

Legend for Dataset S5 (separate file). Phylogenetic tree of *Tas2r* genes in vertebrates. Vertebrate *Tas2r* sequences were obtained from the supplementary data of Li and Zhang (2014), except for those of the 20 frogs examined in this study.

Legend for Dataset S6 (separate file). Ten bitter compounds and their concentrations used in our assays.

SI References

1. H. Jiao, Y. Wang, L. Zhang, P. Jiang, H. Zhao, Lineage-specific duplication and adaptive evolution of bitter taste receptor genes in bats. *Mol. Ecol.* **27**, 4475–4488 (2018).
2. D. Li, J. Zhang, Diet shapes the evolution of the vertebrate bitter taste receptor gene repertoire. *Mol. Bio. Evol.* **31**, 303–309 (2014).
3. A. Krogh, B. Larsson, G. Von Heijne, E. L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
4. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

5. K. Katoh, K. Misawa, K. i. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
6. J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
7. S. Kalyaanamoorthy, B. Q. Minh, T. K. Wong, A. Von Haeseler, L. S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
8. L.-T. Nguyen, H. A. Schmidt, A. Von Haeseler, B. Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
9. D. T. Hoang, O. Chernomor, A. Von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
10. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
11. G.-H. Lin *et al.*, Transcriptome sequencing and phylogenomic resolution within Spalacidae (Rodentia). BMC Genomics **15**, 1–9 (2014).
12. M. Pertea, D. Kim, G. M. Pertea, J. T. Leek, S. L. Salzberg, Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).
13. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
14. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
15. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
16. H. Li *et al.*, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
17. M. Pertea *et al.*, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
18. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
19. C. Chen *et al.*, TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **13**, 1194–1202 (2020).
20. A. Drewnowski, C. Gomez-Carneros, Bitter taste, phytonutrients, and the consumer: a review. *Am. J. Clin. Nutr.* **72**, 1424–1435 (2000).
21. W. Meyerhof, Elucidation of mammalian bitter taste. *Rev. Physiol. Biochem. Pharmacol.* **154**, 37–72 (2005).
22. A. Wiener, M. Shudler, A. Levit, M. Y. Niv, BitterDB: a database of bitter compounds. *Nucleic Acids Res.* **40**, D413–D419 (2012).
23. S. S. Duffey, Sequestration of plant natural products by insects. *Annu. Rev. Entomol.* **25**, 447–477 (1980).
24. S. P. Wooding, V. A. Ramirez, M. Behrens, Bitter taste receptors: Genes, evolution and health. *Evol. Med. Pub. Health* **9**, 431–447 (2021).
25. W. Meyerhof *et al.*, The molecular receptive ranges of human TAS2R bitter taste receptors. *Chem. Senses* **35**, 157–170 (2010).
26. B. Bufe, T. Hofmann, D. Krautwurst, J.-D. Raguse, W. Meyerhof, The human TAS2R16 receptor mediates bitter taste in response to β-glucopyranosides. *Nat. Genet.* **32**, 397–401 (2002).
27. H. Imai *et al.*, Functional diversity of bitter taste receptor TAS2R16 in primates. *Biol. Letters* **8**, 652–656 (2012).