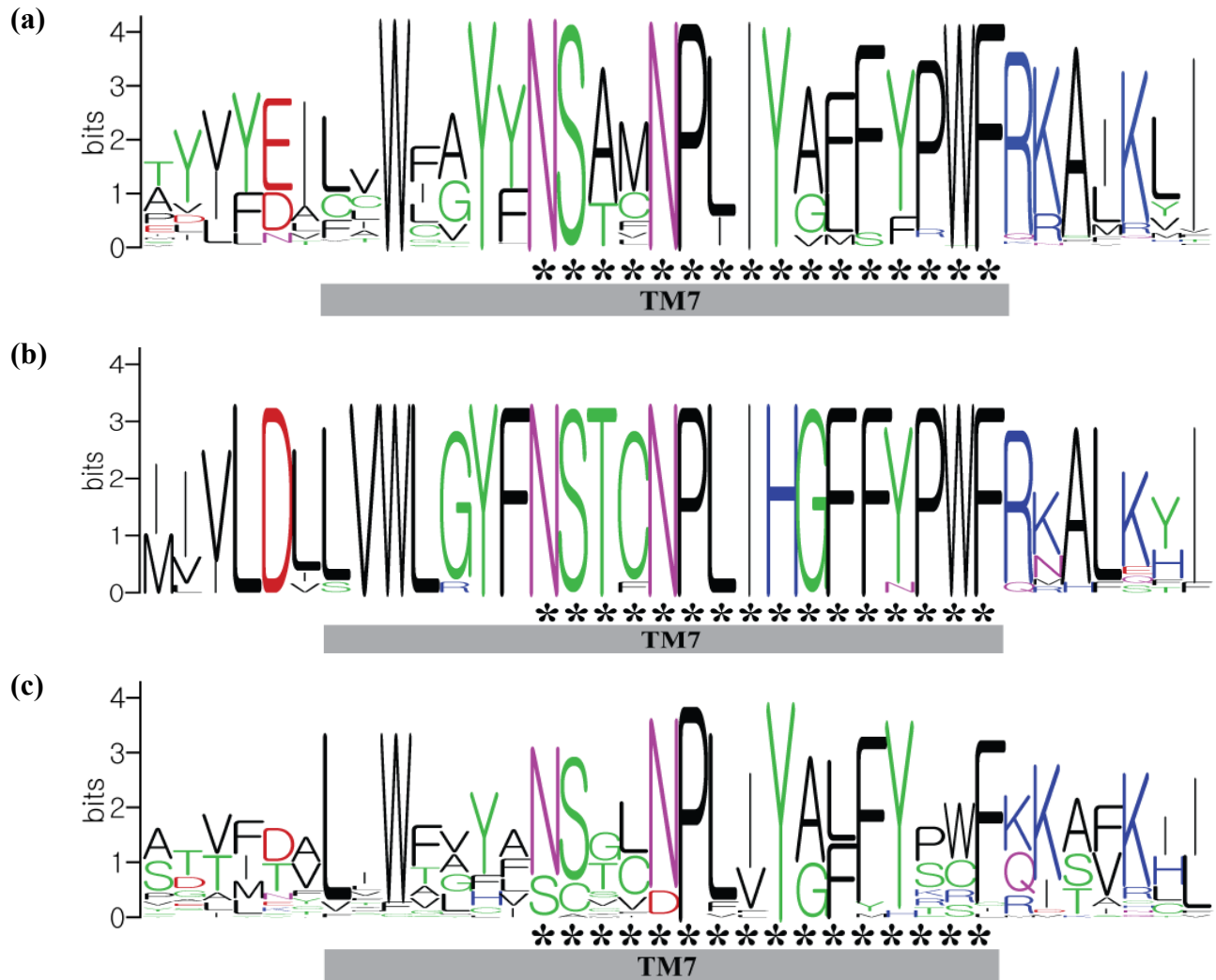


S1 Fig. The maximum-likelihood phylogeny of TAAR proteins from 25 vertebrates. Ten representative biogenic amine receptors (5HT4R: serotonin receptors, H2R: histamine receptors, D5R: dopamine receptors, and ARa2: adrenergic receptors), three cow opsins, and five representative dog olfactory receptors (ORs) are included as the outgroup. The numbers at internal branches show the bootstrap support values (%) for the maximum-likelihood phylogeny and the posterior probability (%) for the Bayesian inference phylogeny. Support values are shown only for the major internal nodes. Three metatherian-specific and one eutherian-specific TAAR groups are indicated as TAAR M1-M3 and TAAR E1, respectively. Teleost fish proteins are indicated with underline. Brown-colored branches indicate the protein lineages where all proteins have weakly conserved TAAR signature motifs (S2 Fig, see Materials and Methods). Two teleost fish clusters colored in gray have TAARs with mixed types of motifs: conserved, weakly conserved, or lost. Note also that the phylogenetic placement of these teleost fish clusters is not resolved. The expanded phylogeny showing all gene names is available on the website: <http://bioinfolab.unl.edu/emlab/TAAR/>



S2 Fig. TAAR signature motifs from TAAR subfamilies (a), from the TAAR3 subfamily (b), and from weakly conserved fish TAARs (c). Conserved amino acid patterns (Lindemann *et al.*, 2005) based on the multiple sequence alignments from positions 291 – 326 (numbering according to the mouse TAAR3: NP_001008429) are shown using the sequence logo (<http://weblogo.berkeley.edu>) (Crooks *et al.*, 2004). 209 sequences from TAAR1-9, M1-M3, and E1 (a) and 13 sequences from TAAR3 (b) were included in each multiple alignment. For fish TAARs, the sequence logo was generated using only 32 sequences where TAAR signature motif (NSX₂NPX₂[Y/H]X₃YXWF) is not conserved. The height of each amino-acid letter is proportional to its frequency of occurrence in a given position. The known TAAR signature motif (NSX₂NPX₂[Y/H]X₃YXWF) corresponds to the positions marked with *. The location of the seventh transmembrane region (indicated as TM7) was predicted using Phobius (Kall *et al.*, 2007).

References:

- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res.* 14:1188-1190.
- Lindemann L, Ebeling M, Kratochwil NA, Bunzow JR, Grandy DK, et al. (2005) Trace amine-associated receptors form structurally and functionally distinct subfamilies of novel G protein-coupled receptors. *Genomics* 85: 372-385.
- Kall L, Krogh A, Sonnhammer ELL. 2007. Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Res.* 35:W429-432.

>shark TAAR S1a

LCYESVNGSCPRAIRSTGVRITLYLLAVLAILVTLFGNMLVIISIAHFKQLHTPTNYLVF
SLAIADFLGCIVMPYSLIRSIESCWYFGILFCKLHTSF~~DLV~~LCAASIIHLCCISVDRYY
AVCDPLKYKTTITVSTVLIMICLSWALSFLVGFVIIIFLELHLIEIKDFYYHEIACFGGCT
LMMGKVCALVYSTISFYFPAFIMVCIYTKIYLVAKKQARTINNLSRKVQPINEGNSIASQ
RSERKAAKTLGIVMGVFILCWSPYFVCDSIEPFIKYSTPPVLFDAFFWVGYL**NSTFNPMI**
YGFYSWERKALKIILTCKIFAPDSSRINLF

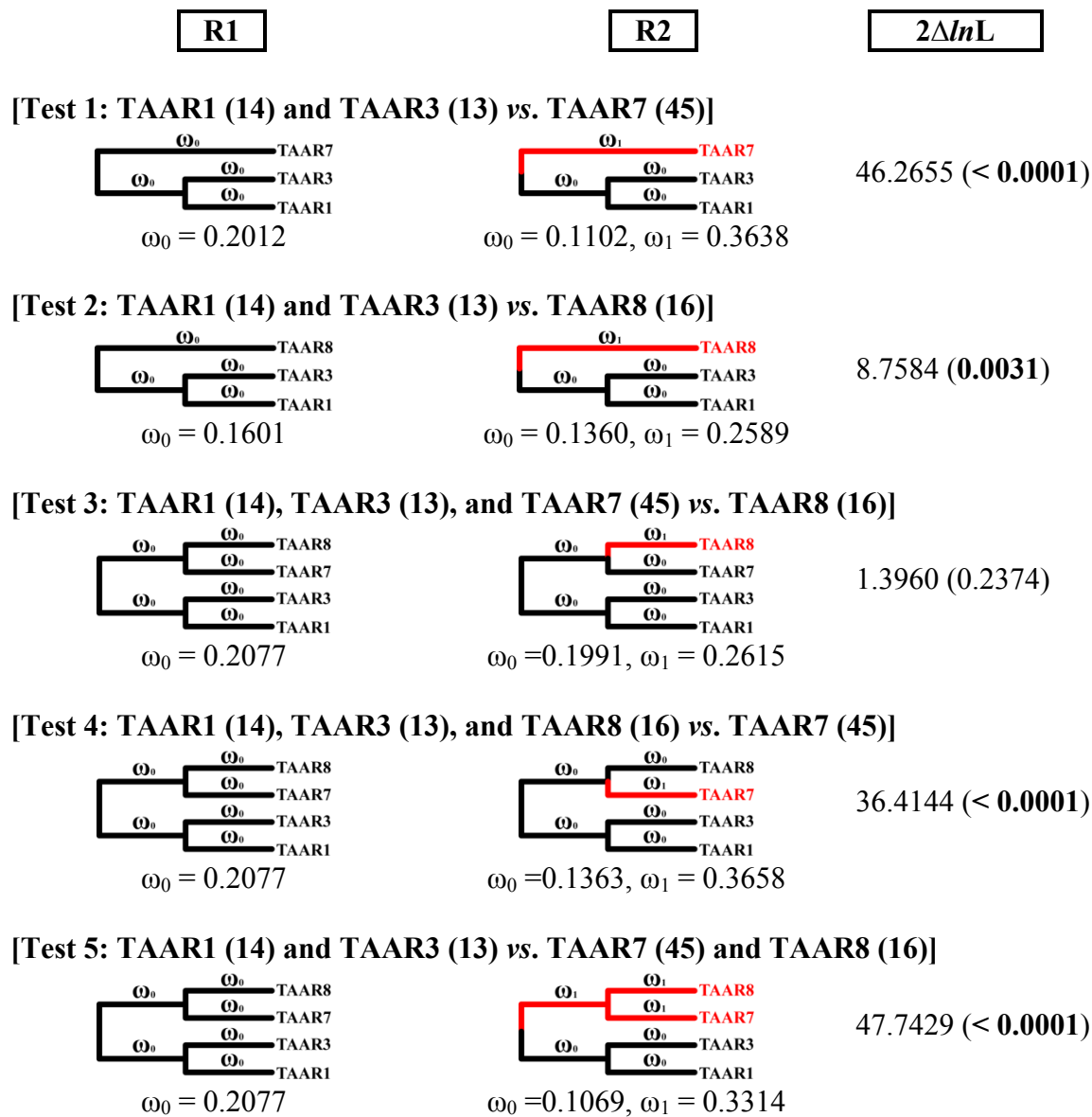
>shark TAAR S2a

MNSINLENSDLQYCFEFNMSCPKSIRSTTTTTVTMYIFITISIVITILGNSVVMISILHF
KQLQTPNTNYLVLSLAFVDFLMGFFVLPFSMVRSVETCWYFGDTFCDIHSTLDVVLTTVSI
YNLCFIAIDRYAVCEPLLYSIKMTLPMTALIIITLNWLF~~AI~~IYGSCVFLSEFTKKASGHY
RTTISCKGSCIEYRFGGHMDALIVLFIPTFIILGIY~~L~~KIYFVQRKHARKIGNMPNNINSK
EEINVRVLQTKKTA~~AK~~NQGVVMGIFVLSWLPFYLSSIINPYLN~~F~~FATPPILFEAFTWFGF
FNSAFNPVLYA~~F~~FYPWERTALKSILTCQILRP~~ESS~~IMNLFPE

S3 Fig. TAAR signature motifs found in the two elephant shark (*Callorhinchus milii*) TAAR protein sequences. The TAAR motif regions are highlighted with yellow. The seven transmembrane regions predicted by Phobius (Kall *et al.*, 2007) are indicated with underline.

Reference:

Kall L, Krogh A, Sonnhammer ELL. 2007. Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Res.* 35:W429-432.



S4 Fig. PAML branch-model tests between primary amine detecting TAARs (TAAR1 and TAAR3) and tertiary amine detecting TAARs (TAAR7 and TAAR8). All tests were performed comparing the two hypotheses: R1 (a single ω for all branches) and R2 (two independent ω 's: ω_1 for the red lineage and ω_0 for the black lineages). The number of the genes included in each TAAR subfamily is given in parentheses after the subfamily name. For the likelihood ratio test statistics, $2\Delta\ln L$, P -values (shown in parentheses) are obtained based on a χ^2 distribution with d.f. = 1. Significant P -values (< 0.05) are shown in boldfaces.

1 . 21 . 41 . 1.30 1.29 1.34

4AMJ MG[]A**ELLSQQWEAGMSLIMALVVLLIVAGNVLVIAAIG**sTQR
 β1AR MGDGWLPPDCGPHNRSGGGGATAAPTGSRQVSAELLSQQWEAGMSLIMALVVLLIVAGNVLVIAAIGRTQR
 humanTAAR1 -----MMPFCHNIINISCVKNNWSNDRASLYSLMVLIIILTTLVGNLIVIVSISHFKQ
 mouseTAAR3 -----MDLIYIPEDLSSCPKFGNKSPPNRSFRVRMIMYLFMTGAMVITIFGNLVI IISISHFKQ
 elephantTAAR7a -----MSTELSPAPVQLCYENLNGSCVKTSYSPGPRVMILYLVFGSGAVLAVFGNLLVMI SMLHFKQ
 mouseTAAR8a -----MTSNFSQPALQLCYENTNGSCKTYPSPGPRVILYMVYGFAGVLAVCGNLLVVISVLHFKQ
 <-----TM1-----><-----

54 61 . 81 . 101 . 121 2.67 3.21

4AMJ LQT**LTNLFITSLACADLVvGLLVVPFGATLVVRGTWLVG**SFLCEI**WTSIDVLCVTAS**IETLCVIAIDRYLA
 β1AR LQTLTNLFITSLACA LVMGLLVVPFGATLVVRGTWLVGSGFLCEI**WTSIDVLCVTAS**IETLCVIAIDRYLA
 humanTAAR1 LHTPTNWLIIHSMATVDFLIGCLIMPYSMVRSABHCWYFGEVFCKIHTSTIMLSASIFHLSFISIDRYA
 mouseTAAR3 LHSPTNFLILSMATDFLLGFVIMPYSMVRSVESCWYFGDSFCKFHASFDMMLSLTSIFHLCSIAIDRFYA
 elephantTAAR7a LHSPANFLIASLACADFLVGVTVMPFSTVRSVESCWYFGEIYCTFHSCFNFSFCYASIFHLCSISVDKFA
 mouseTAAR8a LHSPANFLIASLASADFLVGVISVMPFMSVRSIESCWYFGDAFCSLHSCCDVAFICYSSVHLHCFISVDRYIA
 Δ Δ Δ
 IC1--<-----TM2-----><-----EC1-----><-----TM3----->

125 . 151 . 181 3.65 4.61 5.36

4AMJ **ITSPFRYQSLMTRARAKVICTVWAI**SALVSFLPIMMHWWRDED**PQ-ALKCYQDPGCCDFVTNRA**YAIASS
 β1AR ITSPFRYQSLMTRARAKVICTVWAI**SALVSFLPIMMHWWRDED**PQ-ALKCYQDPGCCDFVTNRA**YAIASS**
 humanTAAR1 VCDPLRYKAKMNILVICVMIFISWSVPAVAFEGMIFLEINFKGABEIIYKHVHCRGGCSVAFSKISGVLVAF
 mouseTAAR3 VCDPLHYTTMTVSMIKRLLAFCAWAPALFSFGLVLSANVSGMQS-YEILVACFNFCALTFNKFWGTILF
 elephantTAAR7a VTDPLIYPTRTASVSGICIAFSWLLSILYSFLLCSGANETGLEE-LVSGLSCVGGCQIAVNSNWVFNVE
 mouseTAAR8a VTDPLVYPTKFTVSVSGICISISWILPLVYSSAVFYTGISAKGIES-LVSALNCVGGCQIVINQDFVLISE
 Δ Δ Δ Δ Δ Δ Δ Δ Δ Δ
 -----IC2-----><-----TM4-----><-----EC2-----><-----

196 201 . 231 . 241 5.65 6.26

4AMJ **IISFYIPLLIMIFVaLRVYREAKEQ**IRKIDR[]ASKRKTS**SRVMIM**
 β1AR **IISFYIPLLIMIFVYLRVYREAKEQ**IRKIDRCEGRFYGSQEQPPPLPQHQPILGNRASKRKTSRVMIM
 humanTAAR1 **MTSFYIPGSIMLCVYRIYLI**AKEQARLISDANQKLQIGL-----EMKNGISQS
 mouseTAAR3 **TTCFFTPGSIMVGIYKIFIVSR**RHARALSMPANTKG-----AVGKNLSKK
 elephantTAAR7a **V-LFFIPTLVMIIVYSKIFLVA**KQQAARKIESLSNKTETSS-----DSYKDRVAK
 mouseTAAR8a **L-LFFIPTLVMIIVYSKIFLVA**KQQAARKIETSVSGNRGESS-----ESHKARVAK
 -----TM5-----><-----IC3----->

245 . 261 . 301 . 6.71 6.78 7.54

4AMJ **REHKALKTLGIIMGVFTLCWLPFFLVNIVNVFN**RDLV**PDWLFVFNWLG**YANSAmNPFIYCR-S**PDFRKA**F
 β1AR **REHKALKTLGIIMGVFTLCWLPFFLVNIVNVFN**RDLV**PDWLFVFNWLG**YANSAmNPFIYCR-S**PDFRKA**F
 humanTAAR1 **KERKAVKTLGIVMGVFLICVCP**FCICVMPFLHYIIPPTINDVLI**FGYLN**STFNPMYAFFYPWFRKAL
 mouseTAAR3 **KDRKAAKTLGIVMGVFLACWLP**CFLAVLIDPYLDYSTPIIVLDLLVWLGYNSTCNPLIHGFYPWFRKAL
 elephantTAAR7a **RRKAAKTLGIAVIAFLISWLP**YFIDIVIDAFLGFITPTIYIYILVWFAYYNSAMNPLIYAFFYPWFRKAI
 mouseTAAR8a **RRKAAKTLGVTVVAFMVSWLP**YITDALVDAFMGFITPAYVYIEICCGTYYSAMNPLIYAFFYPWFKKAI
 -----><-----TM6-----><-----EC3-----><-----TM7----->

316 321 .

4AMJ **KRLLa**FPRKADRRLLhh
 β1AR **KRLLC**FPRKADRRLLHAGGQPAPLPGGFISTLGSPHSPGGTWSDCNGGTRGGSESSLEERHSKTSRSESKM
 humanTAAR1 **KMMLFGKIFQK**DSSRCKLFLELSS-----
 mouseTAAR3 **QFIVSGKIFRS**NSDTANLFPEAH-----
 elephantTAAR7a **KLIITGKVLRE**NSSTNLFS-----
 mouseTAAR8a **KLILSGEILK**GHSSSTANLFSE-----

4AMJ]
 β1AR EREKNILATTRFYCTFLGNGDKAVFCTVLRIVKLFEDATCTCPHTHKLKMKWRFKQHQA
 humanTAAR1 -----
 mouseTAAR3 -----
 elephantTAAR7a -----
 mouseTAAR8a -----

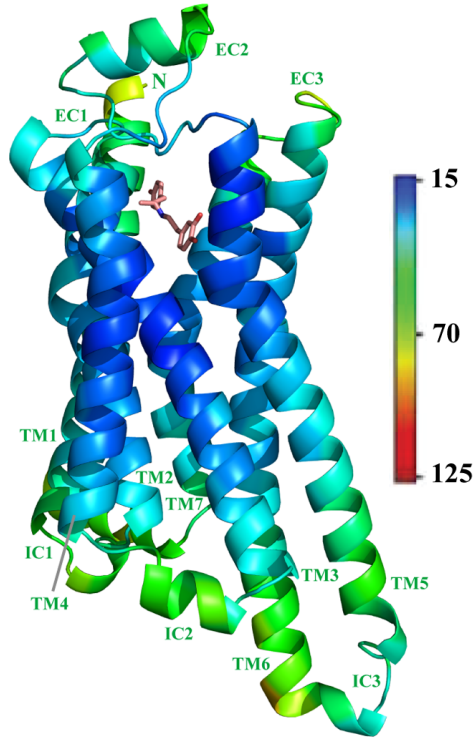
S5 Fig. Multiple alignment of the four TAAR and the turkey β_1 -adrenergic receptor proteins.

Protein sequences of two primary amine detecting TAARs (human TAAR1: NP_612200 and mouse TAAR3: NP_001008429) and two tertiary amine detecting TAARs (elephant TAAR7a: XP_003404143 and TAAR8a: NP_001010830) are aligned with the sequence of the turkey β_1 -adrenergic receptor (β_1 AR: P07700). The position number at the top of the alignment starts at the beginning of the human TAAR1 sequence. Position numbers based on the scheme proposed by Ballesteros and Weinstein (1995) are also shown diagonally for the start and end of each transmembrane region of β_1 AR. Approximate regions for transmembranes (TM1-TM7), intracellular loops (IC1-IC3), and extracellular loops (EC1-EC3) are indicated below each alignment block. The first lines of the alignment show the sequence the protein structure (4AMJ) is based on. In order to improve expression and to obtain crystals, eight thermostabilizing point mutations, a His-tag at the C-terminus, and truncations (at N-terminus, third intracellular loop, and C-terminus) were introduced (Warne *et al.*, 2012). These changes are indicated by lower cases and square brackets in the 4AMJ sequence. Residues assigned for alpha helices in 4AMJ are shown with white letters on black background. 26 residues suggested to involve with agonist binding to the β_1 AR are shown with blue background (Warne *et al.*, 2011, 2012). For the β_1 AR and TAAR protein sequences, residues predicted to be in transmembrane regions by Phobius (Kall *et al.*, 2007) are shown with gray background. The residues surrounding the main and minor ligand-binding pockets in the β_1 AR are shown with cyan and magenta background (Nygaard *et al.*, 2009; Rosenkilde *et al.*, 2010). 29 ligand-binding sites identified by Kleinau *et al.*, (2011) are shown with green background in the human TAAR1. Among them, the residues conserved among human TAARs (including both primary amine detectors and tertiary amine detectors), adrenergic receptors, as well as other biogenic amine receptors are shown with red fonts. Those in the human TAAR1 identical or similar to the residues in the corresponding position of biogenic amine receptors are shown with yellow fonts. Positively selected sites identified by our PAML analysis are shown with triangles below the alignment: red and green are sites identified by the site and branch-site models, respectively, in TAAR7, and purple and brown are sites identified by the site and branch-site models, respectively, in TAAR8. Closed triangles indicate sites identified with posterior probabilities higher than 0.95. See S3 and S4 Tables for details.

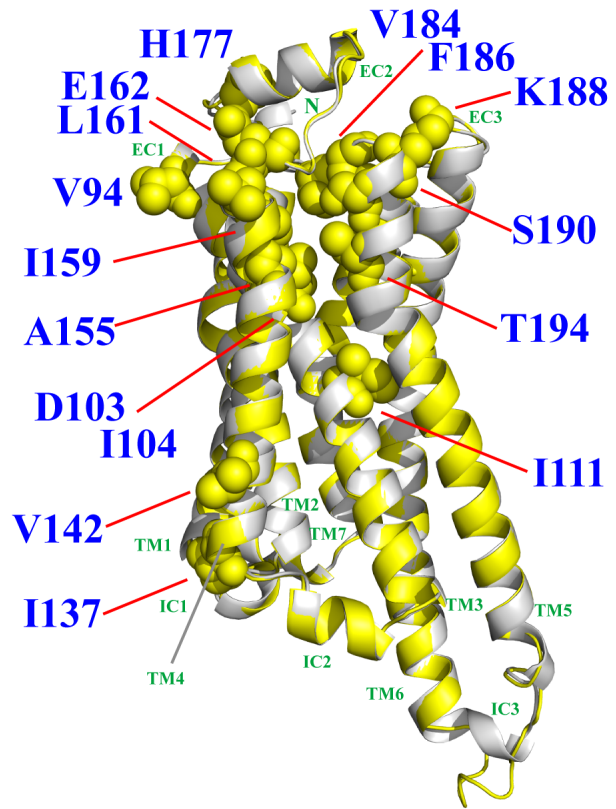
References:

- Ballesteros JA, Weinstein H. 1995. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In: CS Stuart, editor. *Receptor Molecular Biology*: Academic Press. p. 366-428.
- Kall L, Krogh A, Sonnhammer ELL. 2007. Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Res.* 35:W429-432.
- Kleinau G, Pratzka J, Nurnberg D, Gruters A, Fuhrer-Sakel D, Krude H, Kohrle J, Schoneberg T, Biebermann H. 2011. Differential modulation of Beta-adrenergic receptor signaling by trace amine-associated receptor 1 agonists. *PLoS ONE.* 6:e27073.
- Nygaard R, Frimurer TM, Holst B, Rosenkilde MM, Schwartz TW. 2009. Ligand binding and micro-switches in 7TM receptor structures. *Trends Pharmacol Sci.* 30:249-259.
- Rosenkilde MM, Benned-Jensen T, Frimurer TM, Schwartz TW. 2010. The minor binding pocket: a major player in 7TM receptor activation. *Trends Pharmacol Sci.* 31:567-574.
- Warne T, Moukhametzianov R, Baker JG, Nehme R, Edwards PC, Leslie AGW, Schertler GFX, Tate CG. 2011. The structural basis for agonist and partial agonist action on a β_1 -adrenergic receptor. *Nature.* 469:241-244.
- Warne T, Edwards PC, Leslie AG, Tate CG. 2012. Crystal Structures of a Stabilized β_1 -Adrenoceptor Bound to the Biased Agonists Bucindolol and Carvedilol. *Structure.* 20:841-849.

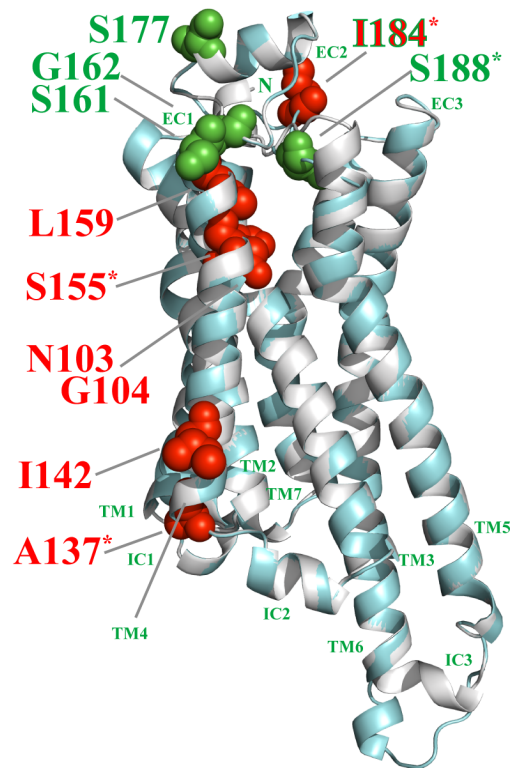
(a)



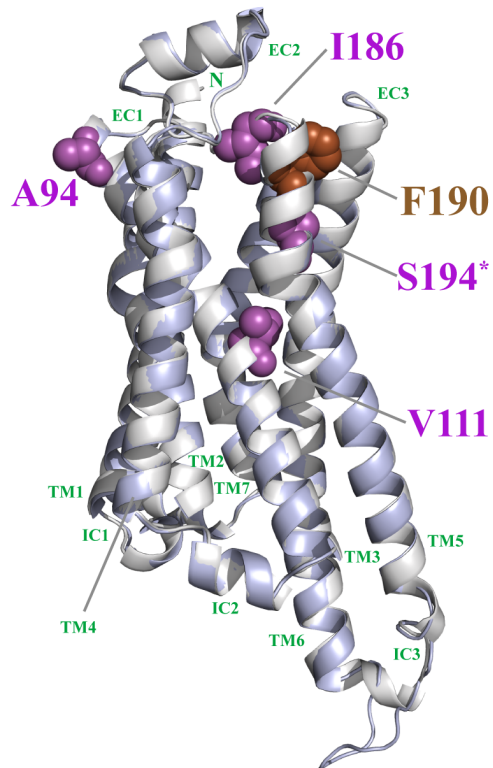
(b)



(c)



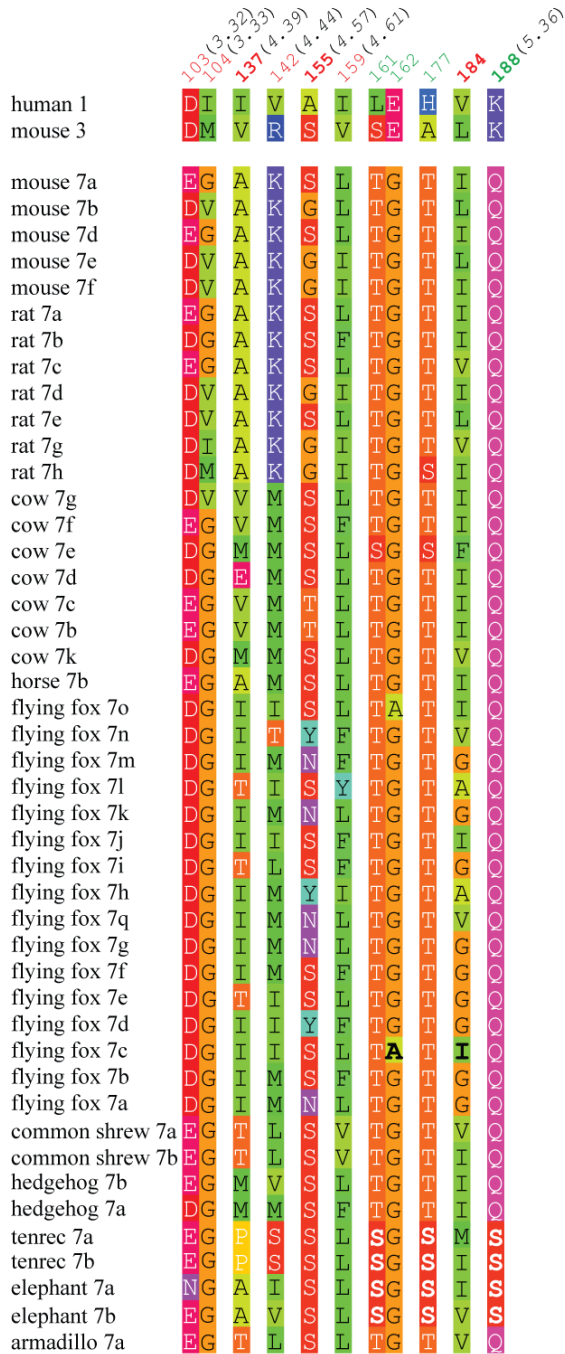
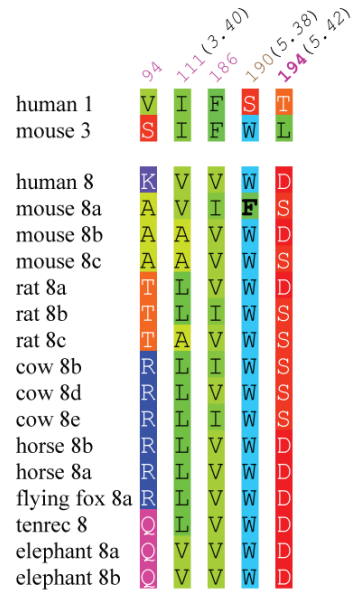
(d)



S6 Fig. Modeling of the 3D-structure of TAAR proteins. The same template, the B-chain of the turkey β_1 -adrenergic receptor (a: β_1 AR, PDB: 4AMJ), was selected by SWISS-MODEL (<http://swissmodel.expasy.org>; Arnold *et al.*, 2006) for modeling protein structures of the human TAAR1 (b: NP_612200), elephant TAAR7a (c: XP_003404143), and mouse TAAR8a (d: NP_001010830) (all E-values < 0.001; their sequence similarities against β_1 AR, P07700, are 49.3%, 46.2%, and 43.9%, respectively). The 3D-structure of the 4AMJ (a) is color-coded based on the temperature factors (B-factors), ranging from 15.74 (blue) to 124.95 (red) (see color scale in the figure). The average B-factor is 45.52. The ligand for the β_1 AR, dobutamine, is shown with the stick model. Note that the template protein contains truncations at N-terminus, third intracellular loop, and C-terminus as well as some thermostabilizing point mutations to improve expression and to obtain crystals (Warne *et al.*, 2012). None of these positions were, however, overlapped with those identified to be under positive selection (see S5 Fig for more details). Predicted protein structures of the human TAAR1 (b: yellow), elephant TAAR7a (c: cyan), and mouse TAAR8a (d: light blue) are superimposed with the template structure (gray) using PyMOL. The QMEAN4 Z-scores given by SWISS-MODEL were -8.27, -8.02, and -8.37 (raw scores: 0.234, 0.250, and 0.228), respectively. The overall root-mean-square deviations (RMSDs) given by PyMOL were 0.054 Å, 0.055 Å, and 0.054 Å, respectively. The N-terminal 15, 25, 23 amino acids (aa) and the C-terminal 19, 16, and 16 aa, respectively, were excluded from the modeling due to insufficient sequence similarity. Positive-selection sites identified by our PAML analysis in elephant TAAR7a (c) and mouse TAAR8a (d) are indicated by red and purple (site models) and by green and brown (branch-site models). Position 184 in elephant TAAR7a was identified by both site and branch-site models. Sites identified with higher than 0.95 posterior probabilities are indicated with asterisks. See S3 and S4 Tables for details on PAML analysis. All amino acid sites corresponding to these positive-selection sites are also mapped on human TAAR1 by yellow spheres for comparison (b). All amino acid position numbers are according to the human TAAR1 sequence. The transmembrane (TM) and internal/external loop (IC1-3 and EC1-3) regions as well as the N-terminal (N) are labeled in each structure. The C-terminal is invisible locating behind TM1. See S5 Fig for the alignment and more detailed information on these sequences.

References:

- Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*. 22:195-201.
- Warne T, Edwards PC, Leslie AG, Tate CG. 2012. Crystal Structures of a Stabilized β_1 -Adrenoceptor Bound to the Biased Agonists Bucindolol and Carvedilol. *Structure*. 20:841-849.

(a)**(b)**

S7 Fig. Alignments of the positively selected sites identified in TAAR7 (a) and TAAR8 (b). The position numbers correspond to those given in S5 Fig. The residues identified by the branch-site models are shown in boldface. The amino acids are color-coded based on their physico-chemical properties using the Taylor color scheme (Taylor 1997). Color-coding is roughly as follows: red for negatively charged (D and E), blue/blueish for positively charged (R, K, and H), green/yellow green for hydrophobic (I, F, V, L, M, and A), blueish green for aromatic (W and Y), purple for large polar (N and Q), and reddish/orange for small (G, T, and S).

References:

Taylor WR. 1997. Residual colours: a proposal for aminochromography. *Protein Eng.* 10:743-746.

S1 Table. The animal genomes used in this study.

Group/species	Order	Sources ^a	Coverage or version	Number of OR genes ^b	Number of TAAR genes ^b
[Euarchontoglires]					
<i>Homo sapiens</i>	Primate	NCBI (BUILD.37.2)	-	388 (414) ^c	6 (3)
<i>Mus musculus</i>	Rodentia	NCBI (BUILD.38.1)	-	1063 (328) ^c	15 (1)
<i>Rattus norvegicus</i>	Rodentia	NCBI (BUILD.4.1)	-	1259 (508) ^c	17 (2)
[Laurasiatheria]					
<i>Bos taurus</i>	Cetartiodactyla	BC	7.1×	970 (1159) ^c	21 (8)
<i>Tursiops truncatus</i>	Cetacea	BI	2.59×	26 ^c	0 (3)
<i>Equus caballus</i>	Perissodactyla	BI	6.79×	NA	11 (4)
<i>Canis familiaris</i>	Carnivora	BI	7.6×	822 (278) ^c	2 (2)
<i>Pteropus vampyrus</i>	Chiroptera	BI	2.63×	672 ^d	26 (10)
<i>Myotis lucifugus</i>	Chiroptera	BI	1.84×	659 ^d	6 (1)
<i>Sorex araneus</i>	Insectivora	BI	1.92×	NA	9 [1] (3)
<i>Erinaceus europaeus</i>	Insectivora	BI	1.86×	NA	6 [2] (4)
[Afrotheria]					
<i>Echinops telfairi</i>	Afrosoricida	BI	1.90×	NA	9 [1] (7)
<i>Loxodonta africana</i>	Proboscidea	BI	1.94×	NA	9 [3] (3)
[Xenarthra]					
<i>Dasylops novemcinctus</i>	Cingulata	WU	2.11×	NA	5 (4)
[Marsupialia]					
<i>Macropus eugenii</i>	Diprotodontia	Ens	2.0×	NA	18 [1] (3)
<i>Monodelphis domestica</i>	Didelphimorphia	BI	6.8×	1198 (294) ^c	22 (4)
[Prototheria]					
<i>Ornithorhynchus anatinus</i>	Monotremata	WU	6.0×	348 (370) ^c	4 (1)
[Sauropsida]					
<i>Gallus gallus</i>	Galliformes	WU	6.6×	211 [89] (133) ^e	4 (1)
<i>Taeniopygia guttata</i>	Passeriformes	WU	6.3×	NA	1 (0)
<i>Anolis carolinensis</i>	Squamata	BI	6.3×	112 [4] (30) ^e	3 (0)
[Amphibia]					
<i>Xenopus tropicalis</i>	Anura	JGI	7.65×	824 [200] (614) ^e	7 (0)
[Teleostei]					
<i>Takifugu rubripes</i>	Tetraodontiformes	IMC	8.7×	47 [39] (39) ^e	18 (1)
<i>Tetraodon nigroviridis</i>	Tetraodontiformes	Gen	8.2×	11 [4] (19) ^e	34 (3)
<i>Danio rerio</i>	Cypriniformes	-	-	154 [1] (21) ^e	110 (10) ^g
[Chondrichthyes]					
<i>Callorhynchus milii</i>	Chimaeriformes	IMC	1.4×	1 [1] (0) ^e	2 (3)
[Agnatha]					
<i>Petromyzon marinus</i>	Petromyzontiformes	UCSC	Ver.2	32 [8] (27) ^e	25 (3)
[Cephalochordata]					
<i>Branchiostoma floridae</i>	Amphioxiformes	JGI	8.1×	31 [3] (9) ^e	0
[Urochordata]					
<i>Ciona intestinalis</i>	Enterogona	JGI	11×	0 (0) ^e	0
<i>Ciona savignyi</i>	Enterogona	ASL (v2.1)	-	0 (0) ^e	0
[Cnidaria]					
<i>Nematostella vectensis</i>	Actiniaria	JGI	7.8×	45 ^f	0

^aData source abbreviations. ASL: the Arend Sidow Lab at Stanford University (<http://mendel.stanford.edu/sidowlab/ciona.html>), BC: Baylor College of Medicine Human Genome Sequencing Center (<http://www.hgsc.bcm.tmc.edu>), BI: Broad Institute at MIT (<http://www.broad.mit.edu>), Ens: Ensembl Genome Browser (<http://www.ensembl.org>), Gen: Genoscope (<http://www.genoscope.cns.fr>), IMC: the Institute of Molecular and Cellular Biology (<http://www.imcb.a-star.edu.sg>), JGI: the Joint Genome Institute (<http://www.jgi.doe.gov>), NCBI: National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>), WU: the Genome Sequencing Center at Washington University School of Medicine (<http://genome.wustl.edu>), and UCSC the University of California-San Cruz (UCSC) Genome Browser (<http://genome.ucsc.edu/>).

^bGene candidates are divided into three categories: intact, incomplete, and pseudogenes. See Table 1 for the details.

^{c-g}The numbers were taken from the following literatures: Nei *et al.* (2008)^c, Hayden *et al.* (2010)^d, Niimura (2009)^e, Churcher and Taylor (2011)^f, and Hashiguchi and Nishida (2007)^g.

NA: not available.

References:

- Churcher AM, Taylor JS. 2011. The Antiquity of Chordate Odorant Receptors Is Revealed by the Discovery of Orthologs in the Cnidarian *Nematostella vectensis*. *Genome Biol Evol.* 3:36-43.
- Hashiguchi Y, Nishida M. 2007. Evolution of Trace Amine-Associated Receptor (TAAR) Gene Family in Vertebrates: Lineage-specific Expansions and Degradations of a Second Class of Vertebrate Chemosensory Receptors Expressed in the Olfactory Epithelium. *Mol Biol Evol.* 24:2099–2107.
- Hayden S, Bekaert M, Crider TA, Mariani S, Murphy WJ, Teeling EC. 2010. Ecological adaptation determines functional mammalian olfactory subgenomes. *Genome Res.* 20:1-9.
- Nei M, Niimura Y, Nozawa M. 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet.* 9:951-963.
- Niimura Y. 2009. On the Origin and Evolution of Vertebrate Olfactory Receptor Genes: Comparative Genome Analysis Among 23 Chordate Species. *Genome Biol Evol.* 2009:34-44.

S3 Table. The results of PAML site-model analysis for TAAR subfamilies.

TAAR subfamily ^a	ω (M0)	$2\Delta\ln L^b$		Positively selected sites ^c
		M2a–M1a	M8–M7	
TAAR1 (14)	0.1807	0 (1)	0.00038 (0.9998)	
TAAR2 (15)	0.0783	0 (1)	0.00408 (0.9980)	
TAAR3 (13)	0.0774	0 (1)	0.00532 (0.9973)	
TAAR4 (15)	0.1406	0 (1)	0.12241 (0.9406)	
TAAR5 (14)	0.1388	0 (1)	3.33783 (0.1885)	
TAAR6 (14)	0.1891	0.2721 (0.8728)	2.1408 (0.3429)	
TAAR7 (45)	0.3512	28.3281 (<0.0001)	36.6892 (<0.0001)	103 ^{5.52} (0.69), 104 ^{5.55} (0.74), 137^{4.39} (0.97) , 142 ^{4.44} (0.89), 155^{4.57} (1.00) , 159 ^{4.61} (0.85), 184 (0.99)
TAAR8 (16)	0.2698	0 (1)	6.84249 (0.03267)	94 (0.59), 111 ^{5.40} (0.78), 186 (0.62), 194^{5.42} (0.95)
TAAR9 (17)	0.1479	0 (1)	0.00024 (0.9999)	
TAAR E1 (6)	0.2835	0 (1)	0.00001 (1)	
TAAR M1 (2)	0.2444	0.0171 (0.9915)	0.06897 (0.9661)	
TAAR M2 (11)	0.3277	1.3045 (0.5209)	5.59743 (0.06089)	
TAAR M3 (9)	0.3102	0 (1)	0.32545 (0.8498)	

^aThe number of the TAAR subfamily genes we tested is given in parentheses.

^bLikelihood-ratio test statistics. *P*-values (shown in parentheses) are obtained based on a χ^2 distribution with d.f. = 2. Significant *P*-values (< 0.05) are shown in boldfaces.

^cPositively selected amino acid sites using the Bayes Empirical Bayes inference with the model M8. The same sites were identified with the model M2a except for two sites (94 and 186). Posterior probabilities are given in parentheses, shown in boldfaces when *P* > 0.95. The position numbers are based on the alignment shown in supplementary S5 Fig. The numbering of the Ballesteros-Weinstein scheme (Ballesteros and Weinstein 1995) is shown in superscripts.

References:

Ballesteros JA, Weinstein H. 1995. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In: CS Stuart, editor. Receptor Molecular Biology: Academic Press. p. 366-428.

S4 Table. The results of PAML branch-site model analysis.^a

TAAR subfamily ^b	Foreground branch	$2\Delta\ln L^c$	Proportion of site class	ω	Positively selected sites ^d
TAAR7 (45)	flying fox TAAR7c	3.9934 (0.0457)	0: 0.68747, 1: 0.29542, 2a: 0.01197, 2b: 0.00514	$\omega_0=0.11593$, $\omega_1=1$, $\omega_2=140.19823$	A162 (0.657), I184 (0.599)
TAAR7 (45)	tenrec- elephant TAAR7	7.2427 (0.0071)	0: 0.69211, 1: 0.29130, 2a: 0.01167, 2b: 0.00491	$\omega_0=0.11524$, $\omega_1=1$, $\omega_2=169.33093$	S161 (0.581), S177 (0.522), S188^{5,36} (0.973)
TAAR8 (16)	mouse TAAR8a	6.0053 (0.0142)	0: 0.82235, 1: 0.17302, 2a: 0.00383, 2b: 0.00081	$\omega_0=0.14625$, $\omega_1=1$, $\omega_2=777.9954$	F190 ^{5,38} (0.935)

^aOnly the results where the given foreground branch having positive selection is supported significantly are listed. These branches are indicated with red color and arrows in Fig 2.

^bThe number of the TAAR subfamily genes tested is given in parentheses.

^cLikelihood-ratio test statistics. *P*-values (shown in parentheses) are obtained based on a χ^2 distribution with d.f. = 1. *P*-values smaller than 0.01 are shown in boldfaces.

^dPositively selected amino acid sites using the Bayes Empirical Bayes inference. Posterior probabilities are shown in parentheses, in boldfaces when *P* > 0.95. The position numbers are based on the alignment in S5 Fig. The numbering of the Ballesteros-Weinstein scheme (Ballesteros and Weinstein 1995) is shown in superscripts.

References:

Ballesteros JA, Weinstein H. 1995. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In: CS Stuart, editor. Receptor Molecular Biology: Academic Press. p. 366-428.