Supporting Information

Zhuang et al. 10.1073/pnas.0808378106

SI Materials and Methods

Primate Genomic DNA Sources. Selected primate genomic DNA was purchased from the Coriell Cell Repositories, including 11 chimpanzees (*Pan troglodytes*, NG06939, NG03487, NG03489, NG03610, NG03622, NG03623, NG03641, NG03650, NG03656, NG03659, and NG03660), 1 bonobo (*Pan paniscus*, NG05253), 1 gorilla (*Gorilla gorilla*, NG05251), 1 Borneo orangutan (*Pongo pygmaeus*, NG05252), 1 white-cheeked gibbon (*Hylobates leucogenys*, PR01038), 1 patas monkey (*Erythrocebus patas*, NG06116), 1 colobus monkey (*Colobus guereza*, PR00980), 1 rhesus macaque (*Macaca mulatta*, NG08305), 1 pigtailed macaque (*Macaca nemestrina*, NG08452), 1 black-handed spider monkey (*Ateles geoffroyi*, NG05352), and 1 squirrel monkey (*Saimiri sciureus*, NG05311).

Mongoose lemur (*Eulemur mongoz*, DUPC4504) whole blood was purchased from Duke University Primate Center. Genomic DNA was extracted using UltraClean BloodSpin kit (MO BIO Laboratories).

Plasmid Construction. The first 20 aa of human rhodopsin (1), or Rho tag (N-MNGTEGPNFYVPFSNATGVVR-C), were inserted between the NheI and EcoRI sites in pCI. All *OR7D4* and *OR7D1* orthologs, synthetic mutants, and reconstructed hypothetical ancestors were then subcloned into this Rho-pCI between the MluI and NotI restriction sites.

OR7D4 synthetic mutants were generated by site-directed mutagenesis through overlap extension PCR from existing clones. Hypothetical ancestors were generated by multiple mutagenesis steps from the closest existing clones. Although the amino acid sequences of ancestral receptors inferred by PAML and those of the actually reconstructed receptors are identical, the exact nucleotide sequences can be different because we did not change nucleotides that do not cause changes in amino acid.

Cloning and Sequencing of *OR7D4* and *OR7D1* Orthologs. *OR7D4* and *OR7D1* orthologs were amplified from various primate genomic DNA. We designed primers based on available sequence data and performed a series of PCR amplifications. ORF-flanking primers were designed based on bushbaby *OR7D4*/*OR7D1* (forward 5′-ATGGAAGCAGAAAACTATACA-3′ and reverse 5′-TCACAGGCAAGAGGCTGT-3′), chimpanzee *OR7D4* (forward 5′-ATGGAAGCAGAAAACCTTAC-3′ and reverse 5′-ATGGAAGCAGAAAACCTTAC-3′), rhesus macaque *OR7D4* (forward 5′-ATGAAAGCAGAAAACCATACA-3′ and reverse 5′-TCATGGACAAGAGGCTG-3′), chimpanzee *OR7D1* (forward 5′-ATGGAAGCAGAAAACCTTAC-3′ and reverse 5′-TCACATTTGCTTAAGGGAC-3′), and rhesus macaque *OR7D1* (forward 5′-ATGAAAGCAGAAAACCATACA-3′ and reverse 5′-TCACATTTGCTTAAGGGACC-3′).

When the above primers did not amplify from a certain species, we also performed PCR amplifications at reduced annealing temperature using degenerate primers (forward 5′-ATGGAAGCAGAAAACYWTACAGAA-3′ and reverse 5′-TCAYRGRCAAGAGKCKGYYCTGCTGAG-3′) designed based on the all cloned *OR7D4* sequences and (forward 5′-ATGRAAGCAGAAAACYWTACAGAA-3′ and reverse 5′-TCABRGRCAAGAGKCKGYYCTGCTRAG-3′) based on all cloned *OR7D4* and *OR7D1* sequences. If all of the above primers did not amplify for a certain species, a pair of internal primers (forward 5′-TTCCTGTCCATGTACCTGGT-3′ and reverse 5′-GAGGCTGCCCTGCTGAGGAG-3′) designed based on conserved sequences on all cloned *OR7D4* was used. For 4 species, squirrel monkey, spider monkey, patas monkey, and gibbon, no amplicon corresponding to a potential *OR7D4*/*OR7D1* gene or pseudogene was obtained using the ORF-flanking primers or degenerate primers or of expected size using the internal primers, supporting the potential absence or significant modification of *OR7D4*/*OR7D1* gene(s) in these species.

In addition to sequencing the cloned ORs, double-stranded PCR products were also sequenced for verification. The PCR products were either gel-purified using MinElute Gel Extraction Kit (Qiagen), or column-purified using Sephadex (GE HealthCare) or Sephacryl (GE HealthCare) and then sequenced with 3130 or 3730 Genetic Analyzers (Applied Biosystems). For *OR7D4*, primate genomic DNAs from chimpanzee, bonobo, gorilla, orangutan, colobus monkey, rhesus macaque, and pigtailed macaque were amplified with KOD polymerase with untranslated region (UTR) primers upstream (5′-TGAGCTGCCACTTGCTGT-3′) and downstream (5′-AGAGCCGGATATTTAAACC-3′) of the *OR7D4* ORFs. For *OR7D1*, genomic DNAs from chimpanzee, bonobo, gorilla, and orangutan were amplified with UTR primers upstream (5′-CCAAGAGTGATATTGAAGAG-3′) and downstream (5′-CAAAGTCGTAGGTGCTTTG-3′) of the *OR7D4* ORFs. The UTR primers were designed based on the chimpanzee sequence. For the rest of the primate orthologs (lemur *OR7D4*/*OR7D1*, rhesus macaque *OR7D1*, and pigtailed macaque *OR7D1*), which the UTR primers did not amplify, double-stranded PCR products amplified using the ORF-flanking primers were sequenced to their entirety.

***OR7D4*/*OR7D1* Gene Gains and Losses in Primates.** When we performed homology searches using the human *OR7D4* nucleotide sequence against the "nr" in the GenBank database, putative *OR7D4*/*OR7D1* partial sequences were found for gorilla (*Gorilla gorilla*), orangutan (*Pongo pygmaeus*), baboon (*Papio papio*), mangabey (*Cercocebus agilis*), langur (*Trachypithecus auratus*), colobus monkey (*Colobus guereza*), Barbary ape (*Macaca sylvanus*), mongoose lemur (*Eulemur mongoz*), brown lemur (*Eulemur fulvus*), and red-bellied lemur (*Eulemur rubriventer*).

Human *OR7D4* is situated on chromosome 19 in a cluster of 7 intact OR genes and 8 OR pseudogenes. One of these pseudogenes, *OR7D1*p, is adjacent and 95% identical to *OR7D4* at the nucleotide level. The closest intact paralog to *OR7D4* in the human genome, OR7D2, is only 70% identical at the amino acid level. Human OR7D2 does not respond to androstenone or androstadienone (2). In the chimpanzee genome, the orthologs of both *OR7D4* and *OR7D1* have intact ORFs.

OR7D4/*OR7D1* orthologs are among the ORs previously sampled from different primate species in several independent studies (3–5). Rouquier et al. (3) performed an OR sequencing analysis from 10 different primate species using OR degenerate primers matching to the conserved amino acid motifs among ORs. Gilad et al. (4) amplified selected ORs from 18 primate species using OR-specific primers and degenerate primers. These studies generated partial sequences of apparent *OR7D4*/*OR7D1* orthologs, either as putative intact genes or as pseudogenes, from various primate species, including several prosimians, Old World Monkeys, and apes. We performed further data-mining in the genome sequences of orangutan (*Pongo pygmaeus*, a Great Ape species), rhesus macaque (*Macaca mulatta*, an Old World Monkey species), marmoset (*Callithrix jacchus*, a New World Monkey species), and bushbaby (*Otolemur garnetti*, a prosimian species). In the case of the rhesus macaque, we found that the OR annotated as *OR7D4* in the database was

actually an apparent pseudogene representing an *OR7D1* ortholog and there was a partial unannotated sequence corresponding to *OR7D4* in the genome assembly. We did not find apparent *OR7D4*/*OR7D1* orthologs in the 6 times coverage genome sequence of marmoset. In the case of bushbaby, we found only one *OR7D4*/*OR7D1* ortholog. We noted that none of these primate orthologs was from New World Monkey species although a significant number of New World Monkeys were included in the studies mentioned earlier.

Sequencing of the our PCR amplification products verified the presence of *OR7D4*/*OR7D1* orthologs in 8 out of the 12 primate species used, namely, chimpanzee, bonobo, gorilla, orangutan, rhesus macaque, pigtailed macaque, colobus monkey, and mongoose lemur (Fig. S2). Among them, 6 species, chimpanzee, bonobo, gorilla, orangutan, rhesus macaque, and pigtailed macaque, appeared to have both *OR7D4* and *OR7D1* orthologs, whereas we could clone only one *OR7D4*/*OR7D1* ortholog from mongoose lemur. This, together with the fact that we found only one *OR7D4*/*OR7D1* ortholog in the genome sequence of another prosimian species, bushbaby, is consistent with the idea that *OR7D4*/*OR7D1* duplication probably occurred after the split between the lemurs and the rest of the primate order. Sequencing confirmed that *OR7D1* is a pseudogene in human and both macaque species. We did not find a colobus monkey *OR7D1* ortholog.

Four species, gibbon, patas monkey, squirrel monkey, and spider monkey did not appear to have the orthologs to *OR7D4*/*OR7D1*. We noted that gibbon, squirrel monkey, and a squirrel monkey congener (*Saimiri boliviensis*) also did not give apparent *OR7D4* orthologs in a previous study (3). We could not exclude the possibility that *OR7D4*/*OR7D1* orthologs actually exist in some of these species. Whole genome sequencing data from these species would provide more accurate answers.

We also noted a divergence of *OR7D4*/*OR7D1* sequences within species. While the mongoose lemur *OR7D4*/*OR7D1*, gorilla *OR7D4*, and bonobo *OR7D4* did not show polymorphisms, all other PCR products had some level of heterozygosity in the 2 chromosomes of the one individual sequenced. For example, rhesus macaque *OR7D4* was represented by an intact allele and a pseudogenized allele caused by a 2-bp deletion (Fig. S2*A*). Additionally, we found that chimpanzee had an allele with a 4-bp insertion, resulting in a frameshifted nonfunctional allele among the chimpanzees we sequenced. Gorilla *OR7D1* had one pseudogenized allele.

Nucleotide and peptide sequences of *OR7D4* from 8 species (colobus monkey, pigtailed macaque, rhesus macaque, orangutan, gorilla, bonobo, chimpanzee, and human), *OR7D1* from 4 species (orangutan, gorilla, bonobo, and chimpanzee), and *OR7D4*/*OR7D1* from mongoose lemur were aligned using ClustalW as implemented in MEGA4 (6). *OR7D1* sequences from 3 species (pigtailed macaque, rhesus macaque, and human) were excluded due to pseudogenizations. Phylogenetic trees for primate *OR7D4*/*OR7D1* was constructed using a neighbor-joining method in MEGA4 with support for branches assessed by bootstrap analyses of 1,000 replicates. This resulted in a phylogeny roughly congruent to the widely accepted primate phylogeny (7), indicating that all sequences isolated by our PCR strategy were truly orthologous.

Maximum Likelihood Analysis of *OR7D4*/*OR7D1*. Maximum likelihood analysis was performed with *OR7D4*/*OR7D1* sequences with CODEML in the PAML4a or 3.15 software package (8). When the individual of a certain species was heterozygous for *OR7D4*/*OR7D1*, the intact, functional, and/or the most common allele was used in PAML analysis.

To detect positive selection, branch-specific models and site-specific models were used. Global ω values were calculated using all 13 *OR7D4*/*OR7D1* sequences (OR7D4+OR7D1 dataset) by a free-ratio model, which allowed ω to vary along different branches. Various likelihood ratio tests using 1-ratio (M0), 2-ratio, and 3-ratio models were used to determine whether ω values were significantly different in the human, Great Ape, and catarrhine lineages from the background lineages and whether these ω values were >1. Twice the difference of the likelihood values of the tree obtained under each model approximately followed a χ^2 distribution and allowed the calculation of the associated P value to accept or reject the null model.

We first obtained a global dN/dS ratio, or ω value, using model M0, which allowed one ω value across all sites and all lineages in the gene, and found no sign of positive selection on *OR7D4* as a whole, as the ω value was less than 1 ($\omega = 0.4$). This was expected since the majority of the amino acid residues in any given protein were subject to functional and structural constraints and were thus likely to be under purifying selection in the course of evolution. Next, we calculated the global dN/dS ratio for each branch on the tree using the free-ratio model, which allowed ω to vary along each branch. We found that ω was more than 1 in 3 *OR7D4* lineages and one *OR7D1* lineage. Subsequent analysis using 1-ratio, 2-ratio, and 3-ratio branch models allowed for a battery of likelihood ratio tests, confirming that the ω value of the Great Ape *OR7D4* branch was indeed greater than that of the rest of the tree, indicating accelerated evolution.

Whole-gene dN/dS ratio is not designed to identify specific domains/residues of positive selection, especially when the rest of the gene is subject to purifying selection (9). To determine whether any codon position was associated with dN/dS, or ω, significantly >1, multiple alignments of all 13 *OR7D4*/*OR7D1* sequences (OR7D4+OR7D1 dataset), as well as a smaller dataset containing only the *OR7D4* sequences from 8 species and the *OR7D4*/*OR7D1* from lemur (OR7D4 only dataset), were fitted to several site-specific models: M0 allowed one ratio across all sites; M1 allowed 2 ω values, one of neutral selection, with $\omega > 1$ disallowed; M2 added one more site class to M1, with $\omega > 1$ allowed; M3 allowed for K classes of sites with discrete ω values, with $\omega > 1$ allowed; M7 fitted the data to a beta distribution, with $\omega > 1$ disallowed; and M8 added one more site class to M7, with $\omega > 1$ allowed. The neutral models M0, M1, and M7 were compared to the positive selection models M3, M2, and M8, respectively. The M3 vs. M0 test was a test of heterogeneity of ω among sites rather than for positive selection. M2, M3, and M8 identified amino acid residues under positive selection with posterior probabilities calculated by both naïve empirical Bayes (NEB) and BEB approach. Any site with a posterior probability of positive selection >0.95 calculated with BEB was considered as positively selected (10,11).

Using these 3 site-specific models for positive selection, along with their nested null models, we carried out additional likelihood ratio tests asking if there was a subset of amino acid residues of OR7D4/OR7D1 under positive selection when the gene was computationally partitioned into 2 or more site classes allowing at least one class for positive selection. All likelihood ratio tests supported positive selection on OR7D4 in the site models. We then used the empirical Bayes calculations under all positive selection site models to identify specific residues that have been subject to positive selection in primates. We found that some residues in OR7D4 appeared to be under positive selection with a high posterior probability. Positively selected sites predicted using different site-specific models (M2 or M8) or datasets (OR7D4 only or OR7D4 and OR7D1 combined) were consistent.

We also performed branch-site model A (12) to test for positive selection on specific sites within OR7D4/OR7D1 along specific branches. However, likelihood ratio test results were not statistically significant ($P = 0.08$ or higher), probably reflecting the low power for identifying specific sites on specific lineages that evolve under positive selection.

Hypothetical ancestors were reconstructed using the marginal reconstruction algorithm with CODEML in PAML4a. Sequences of *OR7D4* from 8 species (colobus monkey, pigtailed macaque, rhesus macaque, orangutan, gorilla, bonobo, chimpanzee, and human), *OR7D1* from 7 species (pigtailed macaque, rhesus macaque, orangutan, gorilla, bonobo, chimpanzee, and human), and *OR7D4*/*OR7D1* from mongoose lemur were used. When the individual of a certain species was heterozygous for *OR7D4*/*OR7D1*, the intact, functional, and/or the most common allele was used. Only nonsynonymous changes were accounted for in the subsequent site-directed mutagenesis for creating ancestors and their corresponding synthetic mutants.

Luciferase Assay and Data Analysis. Rho-tagged ORs were transfected into the Hana3A cell line (13) along with a short form of human RTP1, hRTP1S, which enhances functional expression of the ORs (14). Cells were grown in a 37 °C incubator with 5% CO_2. Dual-Glo Luciferase Assay System (Promega) was used to measure the firefly and Renilla luciferase activities as previously described (13). CRE-luciferase (Stratagene) was used to measure receptor activation. Renilla luciferase driven by a constitutively active SV40 promoter (pRL-SV40; Promega) was used as an internal control for cell viability and transfection efficiency. Hana3A cells were plated on polyD-lysine-coated 96-well plates (BioCoat; Becton Dickinson). Plasmid DNA of the receptor variants and hRTP1S was transfected using Lipofectamine2000. For each 96-well plate, 1 μg of CRE-Luc, 1 μg of pRL-SV40, 5 μg of OR, and 1 μg hRTP1S were transfected. About 24 h posttransfection, the medium was replaced with CD293 chemically defined medium (Gibco) and then the cells were incubated for 30 min at 37 °C and 5% CO_2. The medium was then replaced with odorant solution diluted in CD293 and the cells were incubated for 4 h at 37 °C and 5% CO_2. We followed the manufacturer's protocols for measuring luciferase and *Renilla* luciferase activities. Luminescence was measured using Wallac Victor 1420 plate reader (Perkin-Elmer).

Data were analyzed with Microsoft Excel and GraphPad Prism 4. Normalized luciferase activity was calculated by the formula [luc(N)-luc(lowest)]/[luc(highest)-luc(lowest)], where luc(N) = luminescence of firefly luciferase divided by luminescence of Renilla luciferase of a certain well; luc(lowest) = lowest luminescence of firefly luciferase divided by luminescence of Renilla luciferase of a plate or a set of plates; and luc(highest) = highest luminescence of firefly luciferase divided by luminescence of Renilla luciferase of a plate or a set of plates.

An F test was used to compare the best-fit values of EC_{50} of the dosage response curves between the hypothetical ancestors and each synthetic mutant to assess whether the dose–response curves for androstenone and/or androstadienone of a mutant was significantly different from those of the respective ancestors. We note that the EC_{50} value of a given receptor can vary between experiments, but the relative sensitivity of the receptor variants remains the same. Data from the same experiment were always used when we performed the statistical tests. The F test, as implemented in GraphPad Prism 4, compares the best-fit value of the selected parameter (LogEC50 in this case) between data sets. It fits the data 2 ways and compares the results with an F test. It will first fit each data set independently; it will then do a global fit, finding one shared (among data sets) best-fit value for each parameter selected. The F test gives an F value, degrees of freedom expressed as DFn and DFd, and a P value. DFn is the numerator degrees of freedom, which is the number of parameters (k). DFd is the denominator degrees of freedom, which is the number of variables (n) minus one plus k. The P value represents supporting or rejecting the null hypothesis, that is, LogEC50 is the same for all data sets. For the Great Ape ancestor and the species in the Old World Monkey lineages, we listed the amino acid changes from the catarrhine ancestor and compared the functional changes with the catarrhine ancestor. For the species in the Great Ape lineage, we listed the amino acid change from the Great Ape or hominine ancestor and compared the functional changes with the respective ancestor. Each residue was sorted into 1 of the 3 different functional classes: decrease in function, no significant change in function, or increase in function by significance of the F test. A Bonferroni correction for multiple comparisons was applied to the F test.

1. Krautwurst D, Yau KW, Reed RR (1998) Identification of ligands for olfactory receptors by functional expression of a receptor library. *Cell* 95:917–926.
2. Keller A, Zhuang H, Chi Q, Vosshall LB, Matsunami H (2007) Genetic variation in a human odorant receptor alters odour perception. *Nature* 449:468–472.
3. Rouquier S, Blancher A, Giorgi D (2000) The olfactory receptor gene repertoire in primates and mouse: Evidence for reduction of the functional fraction in primates. *Proc Natl Acad Sci USA* 97:2870–2874.
4. Gilad Y, Man O, Paabo S, Lancet D (2003) Human specific loss of olfactory receptor genes. *Proc Natl Acad Sci USA* 100:3324–3327.
5. Gilad Y, Wiebe V, Przeworski M, Lancet D, Paabo S (2004) Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol* 2:e5.
6. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.
7. Purvis A (1995) A composite estimate of primate phylogeny. *Philos Trans R Soc London B* 348:405–421.
8. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.
9. Hurst LD (2002) The Ka/Ks ratio: Diagnosing the form of sequence evolution. *Trends Genet* 18:486.
10. Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22:1107–1118.
11. Sackton TB, et al. (2007) Dynamic evolution of the innate immune system in Drosophila. *Nat Genet* 39:1461–1468.
12. Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22:2472–2479.
13. Saito H, Kubota M, Roberts RW, Chi Q, Matsunami H (2004) RTP family members induce functional expression of mammalian odorant receptors. *Cell* 119:679–691.
14. Zhuang H, Matsunami H (2007) Synergism of accessory factors in functional expression of mammalian odorant receptors. *J Biol Chem* 282:15284–15293.

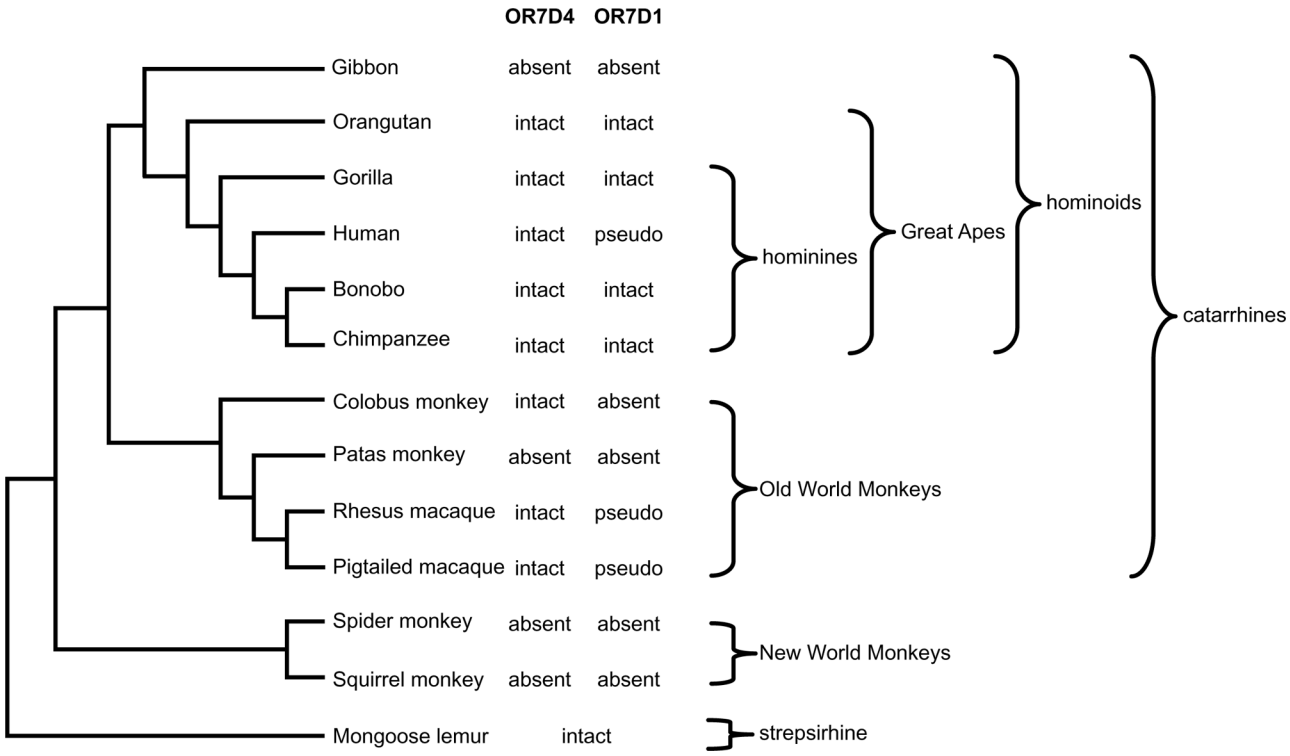

Fig. S1. All primate species whose genomic DNA were used for OR7D4/OR7D1 cloning. Different primate lineage names used throughout the text are indicated by brackets. The status of *OR7D4* and *OR7D1* orthologs for each species is indicated either as intact, pseudo (pseudogene), or absent (not found in the genome).

A

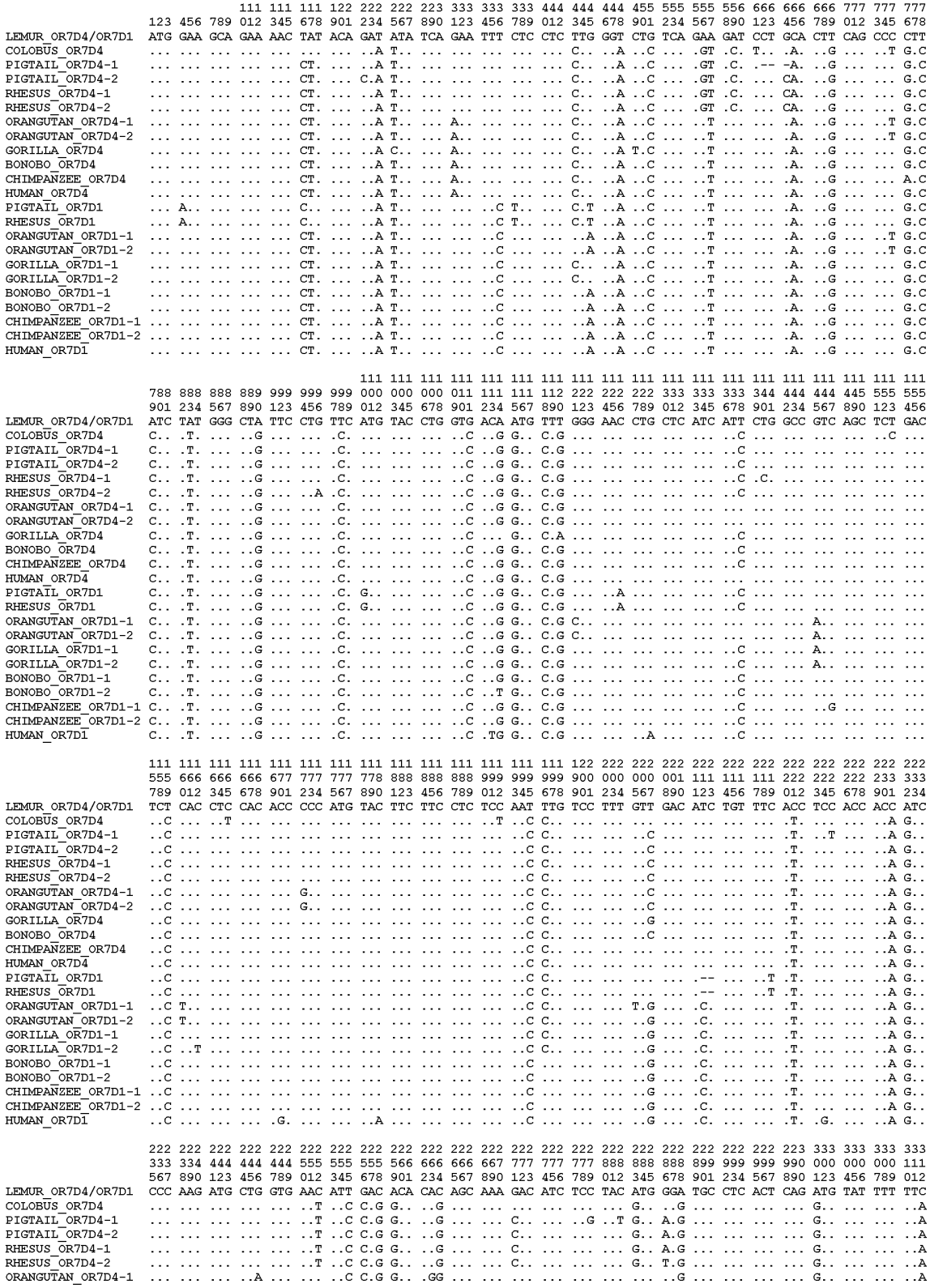

Fig. S2. Various alignments of all *OR7D4*/*OR7D1* alleles cloned. (*A*) Nucleotide alignment of the full ORF of *OR7D4*/*OR7D1* orthologs from 9 primate species. Both alleles of an individual are shown when the species is heterozygous for the gene. (*B*) Peptide alignment of the full ORF of *OR7D4* orthologs from 9 primate species and 7 reconstructed hypothetical ancestors. (*C*) Peptide alignment of the full ORF of *OR7D1* orthologs from 8 species. One allele per species is represented in the peptide alignment in (*B* and *C*). When the individual of a certain species is heterozygous for *OR7D4*/*OR7D1*, the intact, functional, and/or the most common allele is shown in (*B* and *C*).

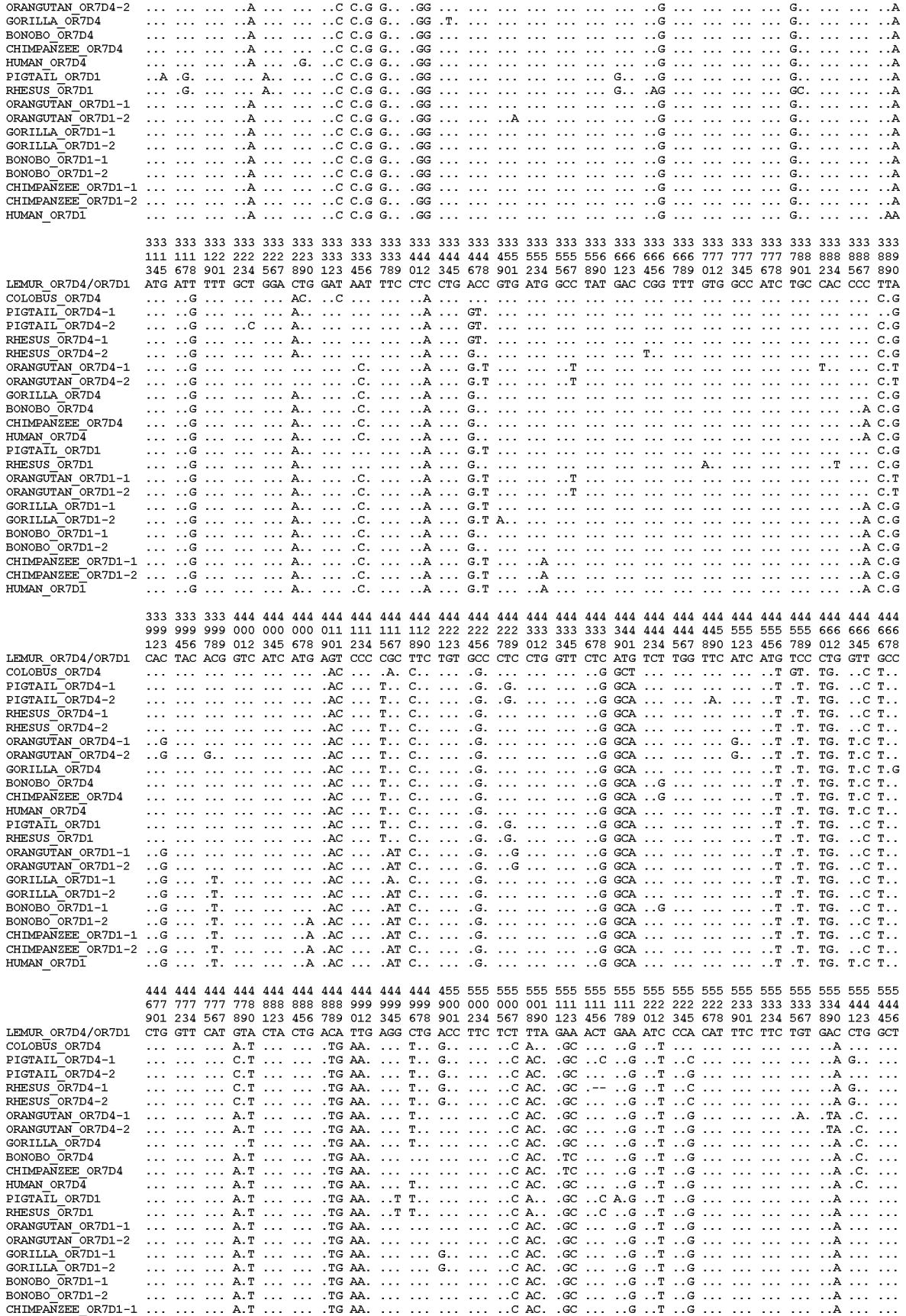

Fig. S2. Continued.

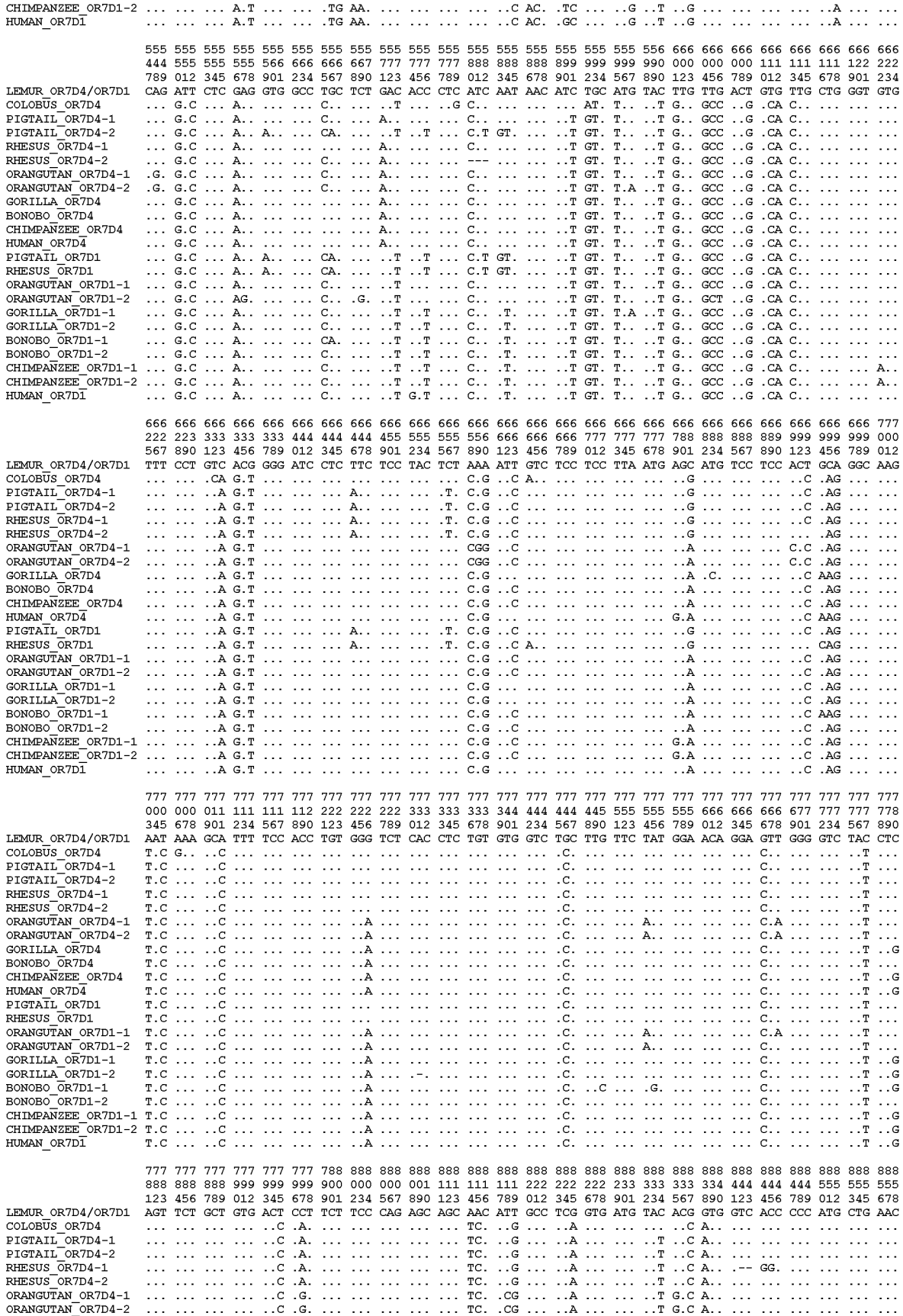

Fig. S2. Continued.

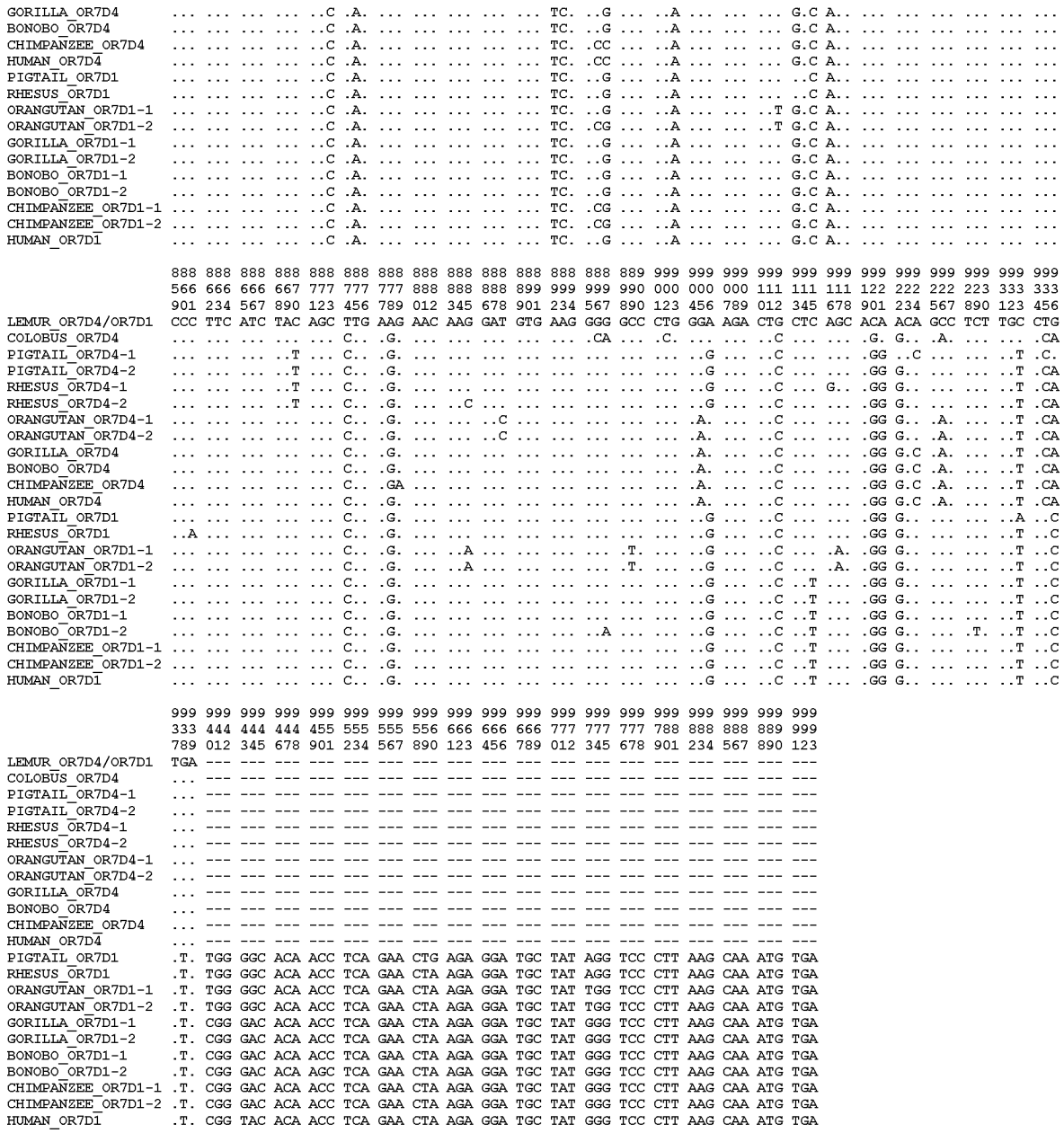

Fig. S2. Continued.

B

```
                          1 1111111112 2222222223 3333333334 4444444445 5555555556 6666666667 7777777778
                 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890
LEMUR_OR7D4/OR7D1 MEAENYTDIS EFLLLGLSED PALQPLIYGL FLFMYLVTMF GNLLIILAVS SDSHLHTPMY FFLSNLSFVD ICFTSTTIPK
CATARRHINE_OR7D4  .....L.EL. ........D. .E...VLF.. ..S.....VL .......... .......... .......... ...I...V..
OWM_OR7D4         .....L.EL. ........GA .E...VLF.. ..S.....VL .......... .......... .......... ...I...V..
COLOBUS_OR7D4     .......EL. ........GA SE...VLF.. ..S.....VL .......... .......... .......... ...I...V..
MACACA_OR7D4      .....L.EL. ........GA .Q...VLF.. ..S.....VL .......... .......... .......... ...I...V..
PIGTAIL_OR7D4-1   .....L.EL. ........GA -Q...VLF.. ..S.....VL .......... .......... .......... ...I...V..
RHESUS_OR7D4-2    .....L.EL. ........GA .Q...VLF.. ..S.....VL .......... .......... .......... ...I...V..
GREAT_APE_OR7D4   .....L.EL. K.......D. .E...VLF.. ..S.....VL .......... .......... .......... ...I...V..
ORANGUTAN_OR7D4-1 .....L.EL. K.......D. .E...VLF.. ..S.....VL .......... .......A.. .......... ...I...V..
HOMININE_OR7D4    .....L.EL. K.......D. .E...VLF.. ..S.....VL .......... .......... .......... ...I...V..
GORILLA_OR7D4     .....L.EL. K.....F.D. .E...VLF.. ..S.....VL .......... .......... .......... ...I...V..
HOMININ_OR7D4     .....L.EL. K.......D. .E...VLF.. ..S.....VL .......... .......... .......... ...I...V..
PAN_OR7D4         .....L.EL. K.......D. .E...VLF.. ..S.....VL .......... .......... .......... ...I...V..
BONOBO_OR7D4      .....L.EL. K.......D. .E...VLF.. ..S.....VL .......... .......... .......... ...I...V..
CHIMPANZEE_OR7D4  .....L.EL. K.......D. .E...ILF.. ..S.....VL .......... .......... .......... ...I...V..
HUMAN_OR7D4       .....L.EL. K.......D. .E...VLF.. ..S.....VL .......... .......... .......... ...I...V..

                          1 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111
                 8888888889 9999999990 0000000001 1111111112 2222222223 3333333334 4444444445 5555555556
                 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890
LEMUR_OR7D4/OR7D1 MLVNIDTHSK DISYMGCLTQ MYFFMIFAGL DNFLLTVMAY DRFVAICHPL HYTVIMSPRF CALLVLMSWF IMSLVALVHV
CATARRHINE_OR7D4  .....QAR.. .......... V..L.M...M .....A.... .......... ......N.CL .G....A... .IFW.S...I
OWM_OR7D4         .....QAQ.. ....V..... V..L.M...M .....A.... .......... ......N.CL .G....A... .IFW.S...I
COLOBUS_OR7D4     .....QAQ.. ....V..... V..L.M...T .......... .......... ......N.HL .G....A... .IVW.S...I
MACACA_OR7D4      .....QAQ.. H...VR.... V..L.M...M .....A.... .......... ......N.CL .G....A... .IFW.S...L
PIGTAIL_OR7D4-1   .....QAQ.. H...VR.... V..L.M...M .....V.... .......... ......N.CL .GR...A... .IFW.S...L
RHESUS_OR7D4-2    .....QAQ.. H...VW.... V..L.M...M .....A.... .W........ ......N.CL .G....A... .IFW.S...L
GREAT_APE_OR7D4   .....QAR.. .......... V..L.M...M .T...A.... .......... ......N.CL .G....A... .IFWFS...I
ORANGUTAN_OR7D4-1 .....QAR.. .......... V..L.M.... .T...A.... .......Y.. Q.....N.CL .G....A... VIFWFS...I
HOMININE_OR7D4    .....QAR.. .......... V..L.M...M .T...A.... .......... ......N.CL .G....A... .IFWFS...I
GORILLA_OR7D4     .....QARI. .......... V..L.M...M .T...A.... .......... ......N.CL .G....A... .IFWFS....
HOMININ_OR7D4     .....QAR.. .......... V..L.M...M .T...A.... .......... ......N.CL .G....A... .IFWFS...I
PAN_OR7D4         .....QAR.. .......... V..L.M...M .T...A.... .......... ......N.CL .G....A... .IFWFS...I
BONOBO_OR7D4      .....QAR.. .......... V..L.M...M .T...A.... .......... ......N.CL .G....A... .IFWFS...I
CHIMPANZEE_OR7D4  .....QAR.. .......... V..L.M...M .T...A.... .......... ......N.CL .G....A... .IFWFS...I
HUMAN_OR7D4       ...S.QAR.. .......... V..L.M...M .T...A.... .......... ......N.CL .G....A... .IFWFS...I

                 1111111111 1111111111 1111111111 1111111112 2222222222 2222222222 2222222222 2222222222
                 6666666667 7777777778 8888888889 9999999990 0000000001 1111111112 2222222223 3333333334
                 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890
LEMUR_OR7D4/OR7D1 LLTLRLTFSL ETEIPHFFCD LAQILEVACS DTLINNICMY LLTVLLGVFP VTGILFSYSK IVSSLMSMSS TAGKNKAFST
CATARRHINE_OR7D4  ..MK.....T G........E ...V.K..R. ...L...VL. VA.A...... .A.......Q ......R... .E..Y.....
OWM_OR7D4         ..MK..A..T G........E ...V.K..R. ...L...VL. VA.A...... .A.......Q ......R... .E..Y.....
COLOBUS_OR7D4     ..MK..A..I G........E ...V.K..R. ...L...IL. VA.A...... AA.......Q .I....R... .E..YE....
MACACA_OR7D4      ..MK..A..T G........E V..V.K..R. N..L...VL. VA.A...... .A...I..FQ ......R... .E..Y.....
PIGTAIL_OR7D4-1   ..MK..A..T G........E V..V.K..R. N..L...VL. VA.A...... .A...I..FQ ......R... .E..Y.....
RHESUS_OR7D4-2    ..MK..A..T G........E V..V.K..R. N..-...VL. VA.A...... .A...I..FQ ......R... .E..Y.....
GREAT_APE_OR7D4   ..MK.....T G........E P..V.K..R. N..L...VL. VA.A...... .A.......Q ......R... .E..Y.....
ORANGUTAN_OR7D4-1 ..MK.....T G.......YV P.RV.K..R. N..L...VL. VA.A...... .A.......R ......R... PE..Y.....
HOMININE_OR7D4    ..MK.....T G........E P..V.K.... N..L...VL. VA.A...... .A.......Q ......R... .E..Y.....
GORILLA_OR7D4     ..MK.....T G........E P..V.K.... N..L...VL. VA.A...... .A.......Q ......RT.. .K..Y.....
HOMININ_OR7D4     ..MK.....T G........E P..V.K.... N..L...VL. VA.A...... .A.......Q ......R... .E..Y.....
PAN_OR7D4         ..MK.....T V........E P..V.K.... N..L...VL. VA.A...... .A.......Q ......R... .E..Y.....
BONOBO_OR7D4      ..MK.....T V........E P..V.K.... N..L...VL. VA.A...... .A.......Q ......R... .E..Y.....
CHIMPANZEE_OR7D4  ..MK.....T V........E P..V.K.... N..L...VL. VA.A...... .A.......Q ......R... .E..Y.....
HUMAN_OR7D4       ..MK.....T G........E P..V.K.... N..L...VL. VA.A...... .A.......Q ......G... .K..Y.....

                 2222222222 2222222222 2222222222 2222222222 2222222222 2222222223 3333333333 333
                 4444444445 5555555556 6666666667 7777777778 8888888889 9999999990 0000000001 111
                 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 123
LEMUR_OR7D4/OR7D1 CGSHLCVVCL FYGTGVGVYL SSAVTPSSQS SNIASVMYTV VTPMLNPFIY SLKNKDVKGA LGRLLSTTAS CL*
CATARRHINE_OR7D4  ........S. .....L.... .....H.... .SM......M .......... ..R....... ......RA.. .P.
OWM_OR7D4         ........S. .....L.... .....H.... .SM......M .......... ..R....... ......RA.. .P.
COLOBUS_OR7D4     ........S. .....L.... .....H.... .SM......M .......... ..R.....A. P.....RAD. .P.
MACACA_OR7D4      ........S. .....L.... .....H.... .SM......M .......... ..R....... ......RA.. .P.
PIGTAIL_OR7D4-1   ........S. .....L.... .....H.... .SM......M .......... ..R....... ......R... .P.
RHESUS_OR7D4-2    ........S. .....L.... .....H.... .SM......M .......... ..R.N..... ......RA.. .P.
GREAT_APE_OR7D4   ........S. .....L.... .....H.... .SM.....AM .......... ..R....... .E....RAD. .P.
ORANGUTAN_OR7D4-1 ........S. .N...L.... .....R.... .ST.....AM .......... ..R....... .E....RAD. .P.
HOMININE_OR7D4    ........S. .....L.... .....H.... .SM.....AM .......... ..R....... .E....RAD. .P.
GORILLA_OR7D4     ........S. .....L.... .....H.... .SM.....AM .......... ..R....... .E....RAD. .P.
HOMININ_OR7D4     ........S. .....L.... .....H.... .SM.....AM .......... ..R....... .E....RAD. .P.
PAN_OR7D4         ........S. .....L.... .....H.... .SM.....AM .......... ..R....... .E....RAD. .P.
BONOBO_OR7D4      ........S. .....L.... .....H.... .SM.....AM .......... ..R....... .E....RAD. .P.
CHIMPANZEE_OR7D4  ........S. .....L.... .....H.... .ST.....AM .......... ..R....... .E....RAD. .P.
HUMAN_OR7D4       ........S. .....L.... .....H.... .ST.....AM .......... ..R....... .E....RAD. .P.
```

Fig. S2. Continued.

C

```
                          1 1111111112 2222222223 3333333334 4444444445 5555555556 6666666667 7777777778
                 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890
LEMUR_OR7D4/OR7D1 MEAENYTDIS EFLLLGLSED PALQPLIYGL FLFMYLVTMF GNLLIILAVS SDSHLHTPMY FFLSNLSFVD ICFTSTTIPK
PIGTAIL_OR7D1    .K...H.EL. ..F.....D. .E...VLF.. ..SV....VL .K........ .......... .......... ?..I...V.R
RHESUS_OR7D1     .K...H.EL. ..F.....D. .E...VLF.. ..SV....VL .K........ .......... .......... ?..I...V.R
ORANGUTAN_OR7D1-1 .....L.EL. ........D. .E...VLF.. ..S.....VL R.......I. ...Y...... ........L. T..I...V..
GORILLA_OR7D1-1  .....L.EL. ........D. .E...VLF.. ..S.....VL ........I. .......... .......... T..I...V..
BONOBO_OR7D1-1   .....L.EL. ........D. .E...VLF.. ..S.....VL .......... .......... .......... T..I...V..
CHIMPANZEE_OR7D1-1 .....L.EL. ........D. .E...VLF.. ..S.....VL .......... .......... .......... T..I...V..
HUMAN_OR7D1      .....L.EL. ........D. .E...VLF.. ..S....MVL .......... ......S..* .......... T..IC..V..

                              1 1111111111 1111111111 1111111111 1111111111 1111111111 1111111111
                 8888888889 9999999990 0000000001 1111111112 2222222223 3333333334 4444444445 5555555556
                 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890
LEMUR_OR7D4/OR7D1 MLVNIDTHSK DISYMGCLTQ MYFFMIFAGL DNFLLTVMAY DRFVAICHPL HYTVIMSPRF CALLVLMSWF IMSLVALVHV
PIGTAIL_OR7D1    ..M..QAR.. ....V..... V..L.M...M .....A.... .......... ......N.CL .GR...A... .IFW.S...I
RHESUS_OR7D1     ..M..QAR.. ....VE.... A..L.M...M .....A.... ...M...... ......N.CL .GR...A... .IFW.S...I
ORANGUTAN_OR7D1-1 .....QAR.. .......... V..L.M...M .T...A.... .......... Q.....N.HL .G....A... .IFW.S...I
GORILLA_OR7D1-1  .....QAR.. .......... V..L.M...M .T...A.... .......... Q.M...N.HL .G....A... .IFW.S...I
BONOBO_OR7D1-1   .....QAR.. .......... V..L.M...M .T...A.... .......... Q.M...N.HL .G....A... .IFW.S...I
CHIMPANZEE_OR7D1-1 .....QAR.. .......... V..L.M...M .T...A.I.. .......... Q.M..IN.HL .G....A... .IFW.S...I
HUMAN_OR7D1      .....QAR.. .......... V..*.M...M .T...A.I.. .......... Q.M..IN.HL .G....A... .IFWFS...I

                 1111111111 1111111111 1111111111 1111111112 2222222222 2222222222 2222222222 2222222222
                 6666666667 7777777778 8888888889 9999999990 0000000001 1111111112 2222222223 3333333334
                 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890
LEMUR_OR7D4/OR7D1 LLTLRLTFSL ETEIPHFFCD LAQILEVACS DTLINNICMY LLTVLLGVFP VTGILFSYSK IVSSLMSMSS TAGKNKAFST
PIGTAIL_OR7D1    ..MKS....I G.K......E ...V.KM.H. ...LV..VL. VA.A...... .A...I..FQ ......R... .E..Y.....
RHESUS_OR7D1     ..MKS....I G........E ...V.KM.H. ...LV..VL. VA.A...... .A...I..FQ .I....R... .Q..Y.....
ORANGUTAN_OR7D1-1 ..MK.....T G........E ...V.K..R. ...L...VL. VA.A...... .A.......Q ......R... .E..Y.....
GORILLA_OR7D1-1  ..MK..A..T G........E ...V.K..R. ...LI..VL. VA.A...... .A.......Q ......R... .E..Y.....
BONOBO_OR7D1-1   ..MK.....T G........E ...V.K..H. ...LI..VL. VA.A...... .A.......Q ......R... .K..Y.....
CHIMPANZEE_OR7D1-1 ..MK.....T G........E ...V.K..R. ...LI..VL. VA.A...M.. .A.......Q ......G... .E..Y.....
HUMAN_OR7D1      ..MK.....T G........E ...V.K..R. .A.LI..VL. VA.A...... .A.......Q ......R... .E..Y.....

                 2222222222 2222222222 2222222222 2222222222 2222222222 2222222223 3333333333 3333333333
                 4444444445 5555555556 6666666667 7777777778 8888888889 9999999990 0000000001 1111111112
                 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890 1234567890
LEMUR_OR7D4/OR7D1 CGSHLCVVCL FYGTGVGVYL SSAVTPSSQS SNIASVMYTV VTPMLNPFIY SLKNKDVKGA LGRLLSTTAS CL*-------
PIGTAIL_OR7D1    ........S. .....L.... .....H.... .SM......M .......... ..R....... ......RA.. *.LWGTTSEL
RHESUS_OR7D1     ........S. .....L.... .....H.... .SM......M .......... ..R....... ......RA.. ..LWGTTSEL
ORANGUTAN_OR7D1-1 ........S. .N...L.... .....H.... .SM.....AM .......... ..R......V .....NRA.. ..LWGTTSEL
GORILLA_OR7D1-1  ........S. .....L.... .....H.... .SM.....AM .......... ..R....... ......RA.. ..LRDTTSEL
BONOBO_OR7D1-1   ........SF .C...L.... .....H.... .SM.....AM .......... ..R....... ......RA.. ..LRDTTSEL
CHIMPANZEE_OR7D1-1 ........S. .....L.... .....H.... .ST.....AM .......... ..R....... ......RA.. ..LRDTTSEL
HUMAN_OR7D1      ........S. .....L.... .....H.... .SM.....AM .......... ..R....... ......RA.. ..LRYTTSEL

                 3333333333 3
                 2222222223 3
                 1234567890 1
LEMUR_OR7D4/OR7D1 ---------- -
PIGTAIL_OR7D1    RGCYRSLKQM *
RHESUS_OR7D1     RGCYRSLKQM *
ORANGUTAN_OR7D1-1 RGCYWSLKQM *
GORILLA_OR7D1-1  RGCYGSLKQM *
BONOBO_OR7D1-1   RGCYGSLKQM *
CHIMPANZEE_OR7D1-1 RGCYGSLKQM *
HUMAN_OR7D1      RGCYGSLKQM *
```

Fig. S2. Continued.

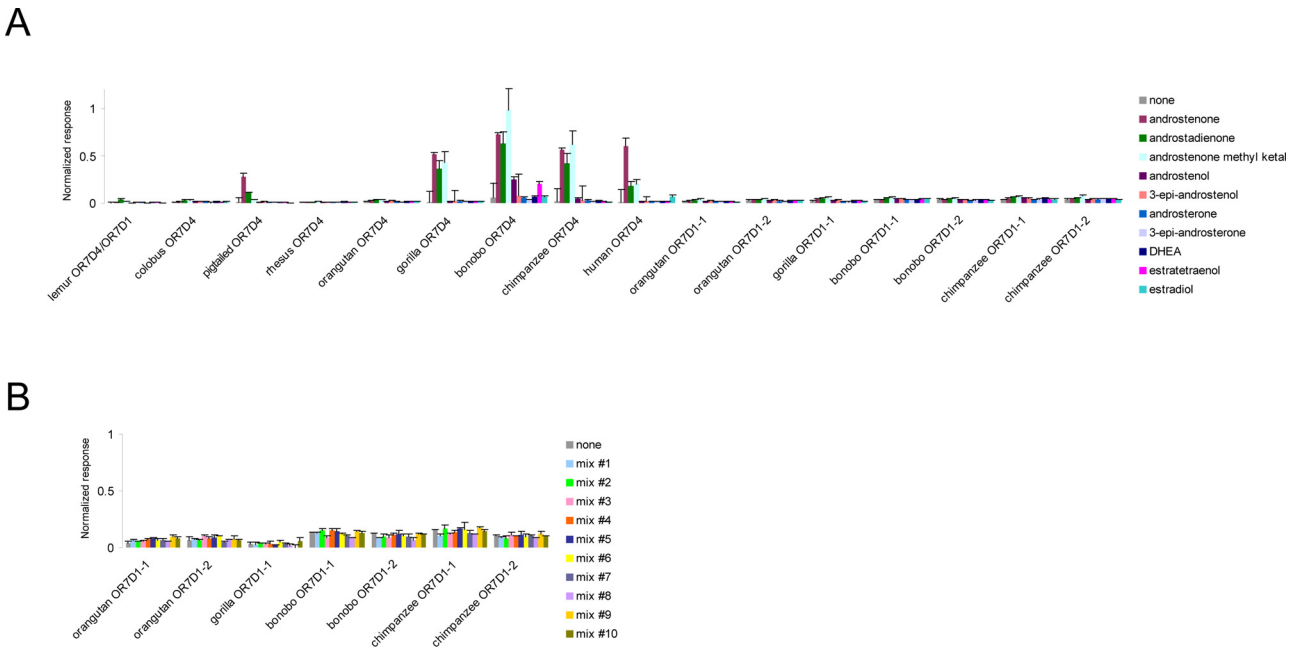

Fig. S3. Primate OR7D1 orthologs do not have a cognate ligand among the odorants tested. (*A*) Screen of primate OR7D4/OR7D1 orthologs against 10 steroids. One allele of OR7D4 ortholog per species and all intact alleles of OR7D1 are shown. (*B*) Screen of primate OR7D1 orthologs against 10 mixtures of odorants. All intact alleles of OR7D1 orthologs are shown. See Table S1 for a list of odorants in each mixture. In (*A* and *B*), *y* axis denotes normalized response ± SEM. (n = 2). In (*B*), *y* axis is normalized to a positive control.

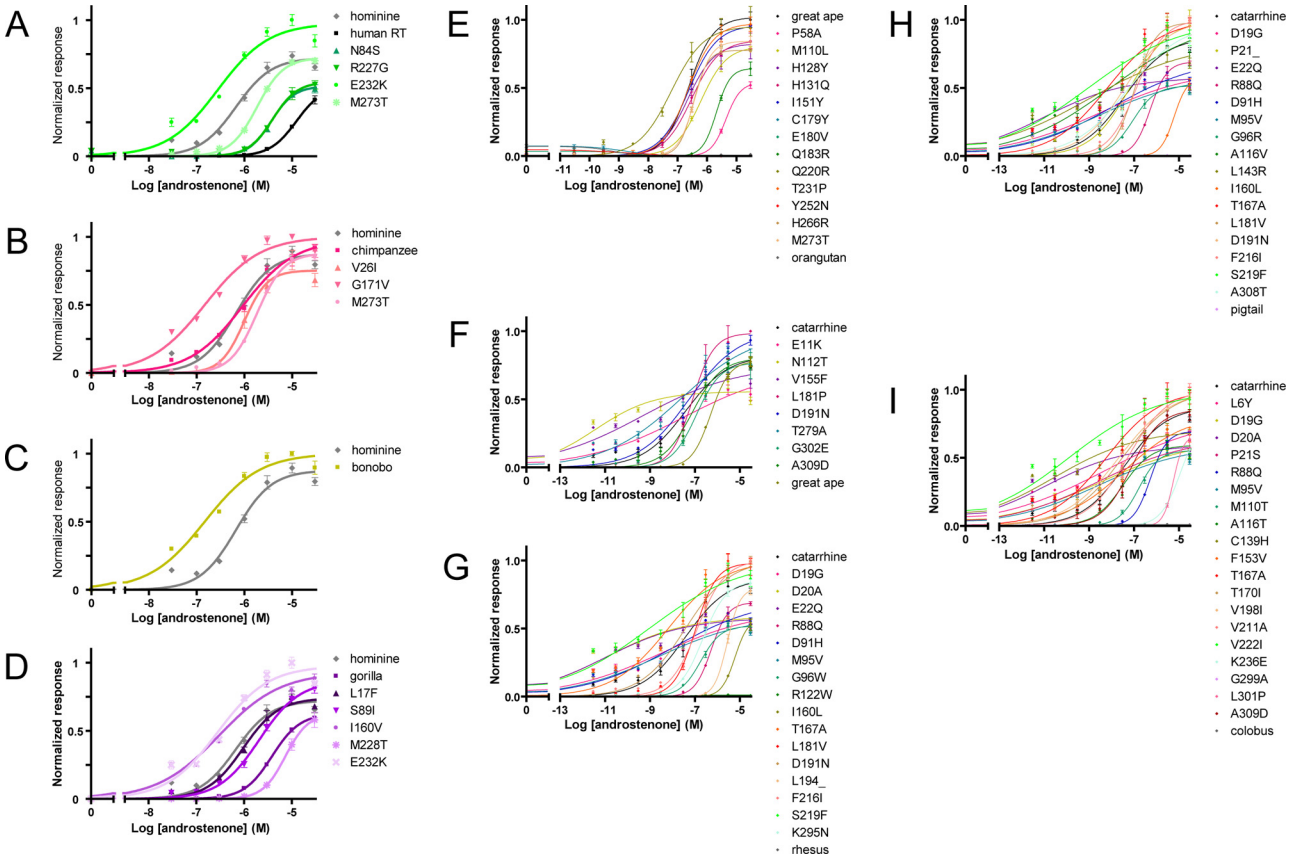

Fig. S4. Point mutations are responsible for the functional divergence of OR7D4 orthologs in response to androstenone. Dose-response curves of the hypothetical ancestor OR7D4 and their respective more derived species/ancestors are shown along with the corresponding synthetic mutants in response to androstenone: (*A*) hominine ancestor and human, (*B*) hominine ancestor and chimpanzee, (*C*) hominine ancestor and bonobo, (*D*) hominine ancestor and gorilla, (*E*) Great Ape ancestor and orangutan, (*F*) catarrhine ancestor and Great Ape ancestor, (*G*) catarrhine ancestor and rhesus macaque, (*H*) catarrhine ancestor and pigtailed macaque, and (*I*) catarrhine ancestor and colobus monkey. *y* axis denotes normalized response ± SEM. (n = 3–4).

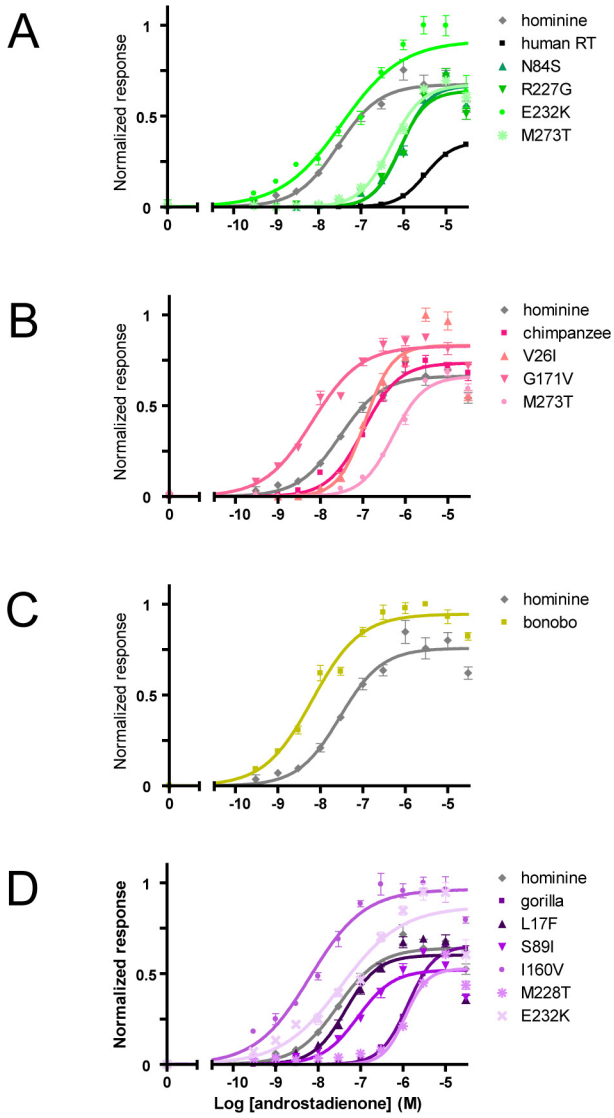

Fig. S5. Selected synthetic mutations are responsible for the functional divergence of OR7D4 orthologs in response to androstadienone. Dose-response curves of the hypothetical hominine ancestor OR7D4 shown with OR7D4 orthologs of (*A*) human, (*B*) chimpanzee, (*C*) bonobo, and (*D*) gorilla and their corresponding synthetic mutants in response to androstadienone. *y* axis denotes normalized response ± SEM. ($n = 4$).

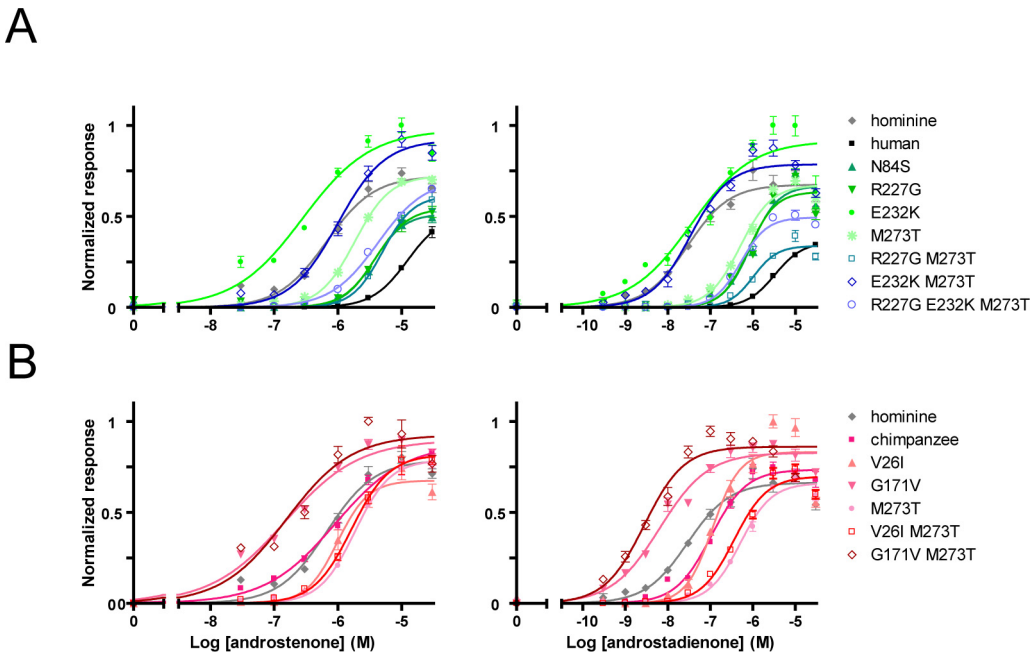

Fig. S6. Variable functional activation of selected hominine double and triple synthetic mutants. Dose-response curves of additional double and triple mutants for simulations of intermediate states between the hypothetical hominine ancestor and (*A*) human and (*B*) chimpanzee to androstenone (*Left*) and androstadienone (*Right*) are shown with all synthetic mutants in these species. *y* axis denotes normalized response ± SEM. ($n = 4$).

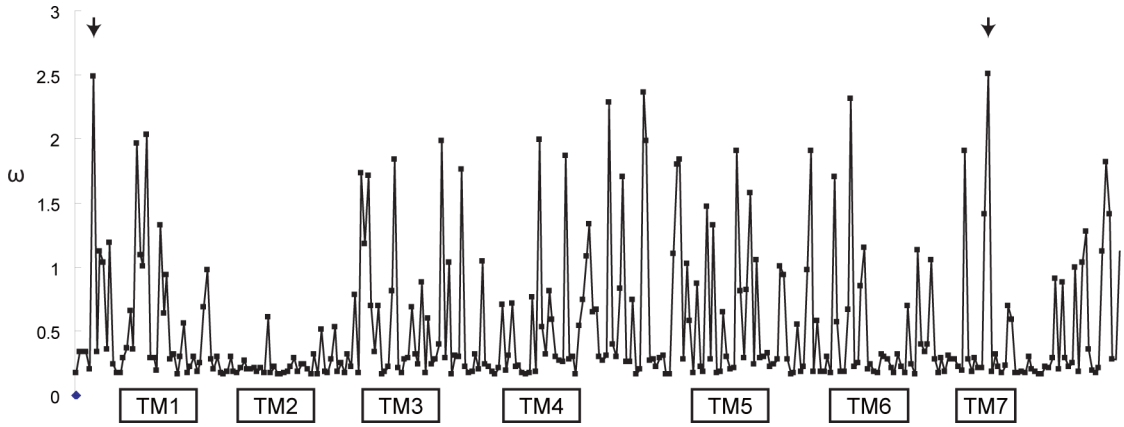

Fig. S7. Variation in ω in the OR7D4 gene. The ω value for each codon is calculated by the M8 positive selection model of PAML using the OR7D4 only dataset shown in Table S5. The approximate limits of predicted transmembrane domains of OR7D4 are represented in boxes. Arrows indicate codons with a posterior probability of positive selection $>$0.95 using the Bayes empirical Bayes calculation.

Other Supporting Information Files

Table 1 (PDF)
Table 2 (PDF)
Table 3 (PDF)
Table 4 (PDF)
Table 5 (PDF)