bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genetic and functional odorant receptor variation in the Homo lineage

Claire A. de March [1,*], Hiroaki Matsunami [1,2] Masashi Abe [1,3], Matthew Cobb [4], Kara C. Hoover [5,*]

[1] Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27710, USA

[2] Department of Neurobiology, Duke Institute for Brain Sciences, Duke University, Durham, NC 27710, USA

[3] Department of Biotechnology and Life Science, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan

[4] Faculty of Life Sciences, The University of Manchester, Oxford Road, Manchester, M13 9PL, USA

[5] Department of Anthropology, University of Alaska Fairbanks, Fairbanks Alaska, 99775, USA

* equal contribution / correspondence

Abstract

The largest and rapidly evolving gene family of odorant receptors detect odors to variable degrees due to amino acid sequence and protein structure. Hybridization between humans, Neandertals, and Denisovans implies shared behavior[1,2], although some speculate that Neandertals were poor smellers [3,4]. We identified genetic and functional variation in humans and extinct lineages in 30 receptors with known function. We show that structural changes in receptor proteins altered odor sensitivity not specificity, indicating a common repertoire across

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

lineages. In humans, variation in receptors may change odor perception or induce odor-specific anosmia [5,6]. Variation in sensitivity may reflect local adaptations (e.g., Denisovan sensitivity to honey, Neandertals sensitivity to grass and sulphur). Extinct human lineages had highly conserved receptor genes and proteins. We observe a similar pattern in the Neandertal OR5P3 variant, which produced no response to ~350 odors. Our data suggest that receptor structure was highly conserved in our closest relatives, but not in living humans. The diversity of geographic adaptations in humans may have produced greater functional variation, increasing our olfactory repertoire and expanding our adaptive capacity [5]. Our results provide insight into odorant receptor function and shed light on the olfactory ecology of ancient humans and their extinct relatives. By studying the function of ancient odorant receptor genes, we have been able to get a glimpse of the sensory world of our extinct ancestors and relatives, with some of the variants giving specific insights into potential adaptations shown by these long-dead populations. The functional variability we have identified in the molecular structure of the odorant receptor proteins will aid in the more general problem of understanding the function of odorant receptor proteins and the neurons they are carried by, opening the road to linking receptor function to perception.

Main

Terrestrial animals smell by binding odorant molecules to odorant receptors (ORs). Variation in mammalian ORs is linked to dietary niche, habitat, and sociability [7].

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mammalian olfactory sensory neurons are biased towards food odors [8] and humans have undergone major evolutionary dietary shifts associated with changes in OR function (e.g. scavenging and hunting, animal milk consumption, cooking, domesticated grains) [7,9-13]. High allelic diversity in human ORs has increased the effective range of detectable odors (olfactory repertoire) [5]. Most human ORs respond to multiple odorants but there is limited knowledge of odorant-OR and genotype-phenotype associations.

The genus *Homo* is unique among primates in its migration and adaption to diverse global environments [1]. Denisovans and Neandertals ancestors dispersed from Africa earlier than humans [2] (~750,000 versus 65,000 years ago) and separated from each other ~300,000 years ago [14] (Extended Data Fig. 1). Neandertals were geographically wide-ranging (western Europe, Middle East, Asia) while Denisovans were geographically constrained to Siberia [3,15], the Tibetan plateau [16], and possibly beyond Wallace's Line [17]. Olfactory stimuli from divergent environments following independent dispersals from Africa may have left traces of variation in *Homo* ORs. Using published ancient DNA sequences to analyze genetic variation for population structure and test functional differences in gene variants for ecological differentiation, we address the question of whether there is a shared *Homo* olfactory repertoire.

We previously explored genetic and functional variation in humans, Altai Neandertal, and Denisovan for OR7D4 [18]. We extended our study to 29 additional open reading frames (Supplemental Information Table 1) for ORs with known human receptor-odor [6,19] and two additional Neandertals (Chagyrskaya, Vindija) and one ancient human

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Ust'-Ishim) who lived in the same Altai montane locality (full sample in Extended Data Fig. 2). We used 1000 Genomes for living humans (Extended Data Fig. 3) [3]. We found that novel variants alter sensitivity but not specificity of OR function and conclude that the olfactory repertoires of extinct lineages were highly overlapping.

Results

Genetic variation. Extinct lineages and Ust'-Ishim had fewer DNA variants than 1000 Genomes (Extended Data Fig. 4). Extinct lineages had smaller effective population sizes and lived in small, isolated communities prone to allele fixation through genetic drift [20]. Several variants were shared among samples (Supplemental Information Table 2), suggesting these variants occurred prior to global divergence of *Homo* (e.g., 6P1, 4Q3, 10G4, and 10G7 in Fig. 1).

The fixation index (Fst) measures genetic variance due to population structure (typically weighted by population size) and ranges from 0 (no differentiation) to 1 (complete differentiation) [21]. 1000 Genomes fixation is lower than other large bodied mammals with wide geographic dispersal (Extended Data Fig. 5) [22]. OR5P3 fixation was the highest at 11.4%, the lower limit of possible significant differentiation by population structure. Comparing 1000 Genomes to extinct lineages, Fst values greater than 20% were observed in 1C1, 5AN1, 7C1, 10G3, 10J5, 51E1, 51E2, indicating past populations were more highly structured than today or that population structure differentiation may be an artifact of the greater variation in 1000 Genomes for these ORs. The remaining Fst values were <1%, indicating either panmixia or fixation due to

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

small group size, but may also these ORs having less variation compared to 1000 genomes or sharing more variants across populations than the high Fst ORs.

A total of 63 DNA variants were observed across 30 OR genes in extinct lineages but 78% of these were shared with 1000 Genomes (Supplemental Information Table 2). Variants observed only in ancient samples made up 22% (n=14) of the total observed variation (Extended Data Fig. 6). Denisovan contained the most novel variants—nine compared to the Neandertal five. No novel variants were observed in the ancient human Ust'-Ishim. Ten out of the 14 novel variants were missense and distributed across eight genes.

Amino acid sequences for the ORs studied were used to form a cladogram (Extended Data Fig. 7). Extinct lineages formed a clade with Vindija Neandertal the most distinct, which was unexpected because genome-wide studies have indicated that Vindija is most genetically similar to Chagyrskaya Neandertal [20].The extinct clade was closest to the ancient human Ust'-Ishim and then to East and South Asian humans—the latter groups harbor genetic signatures of introgression with Neandertals and Denisovans [17].

Functional Variation. Because gene function in not reliably predictable for ORs from sequence data [23,24], we directly measured the functional responses of two Neandertal and six Denisovan ORs containing novel variants. Each OR protein, expressed in a cell line, was screened against seven odorants previously identified in the literature as evoking responses: OR1A1 [23,25], OR1C1 [24], OR2C1 [23], OR10J5 [23], OR5P3 [23], and

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

OR10G3 [24]. Dose response assays for the top screening responses included seven concentrations of the odors delivered separately.

There were only three Neandertal genes containing novel variants. Their dose reponses were not correlated with those of humans (Figs. 2a, 2c, 2d). Only 1C1 had a detectable response but significantly lower than that of humans (Figs. 3a, 3c, 3d, 4, S2, S3). Despite the higher number of novel OR variants in the Denisovans and higher dose responses compared to humans (Fig. 2c, 2d), these OR responses and those for human reference were significantly correlated (R^2 = 0.87) (Fig. 2b). Tested Denisovan ORs were less sensitive to odors we perceive as floral but much more sensitive to odors we perceive as spicy, balsamic, or unpleasant (e.g., sulfur 4x greater than in humans and three times greater for balsamic) (Table 1). Higher dose responses in Denisovan ORs appear to be driven by two amino acid variations in two of the ORs with novel variants (Fig. 3, Extended Data Fig. 8).

Despite using the same DNA sequence as previous studies of 2B11 [24] and 6P1 [6], the human reference for both of these did not respond to any of the ~350 odors (100µM) against which they were tested (Extended Data Fig. 9)—neither did the extinct linages. In previous studies, OR2B11 [24] and OR6P1 [6] responded strongly to coumarin and anisaldehyde, respectively, at concentrations higher than 100µM. We found that such concentrations often cause cell toxicity or OR non-specific cell responses.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

OR1A1. All but 2-heptanone induced a response in the screening assay (Extended Data Fig. 8a) and the highest responses were for honey (particularly for the Denisovan) (Table 1, Fig. 3). Fst between 1000 Genomes and extinct lineages for this gene was low (7%), suggesting some difference due to population structure (1000 Genomes Fst = 3.4%) (Extended Data Fig. 5). Neither of the two Denisovan variants (V461I$^{1.54}$, T257M$^{6.54}$) were in an amino acid region critical for mammalian OR function nor were they involved in the odorant binding cavity—perhaps explaining their minimal functional impact.

OR1C1. The only significant response was from Chagyrskaya Neandertal to androstadienone (Extended Data Fig. 8b) and it was weak (Table 1; Fig. 3). Fst between 1000 Genomes and extinct lineages for 1C1 was high (38%) (1000 Genomes Fst = 9%), suggesting some structure in human groups across evolutionary time (Extended Data Fig. 5). Chagyrskaya Neandertal 1C1 variant Y120H$^{3.48}$ is part of the highly conserved MAY$^{3.48}$DRY motif involved in the activation of mammalian ORs (Fig. 3, Extended Data Fig. 8b), which might explain why this variant alters function.

OR2C1. Screening assay responses were strong but not statistically significant (Extended Data Fig. 8a). The dose response assay for octanethiol produced a statistically significant response in the Denisovan version of this OR (Table 1, Fig. 3). Fst between 1000 Genomes and extinct lineages for this gene was 13.3% compared to that within 1000 Genomes (1.5%), suggesting that human populations are not highly differentiated by this gene today but were to a greater degree in the past (Extended

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Data Fig. 5). The shared C149W$^{4.50}$ corresponds to the conserved W$^{4.50}$. The W residue is highly conserved in GPCRs but less so in ORs (58%). The location of the novel Denisovan I214T$^{5.54}$ in TM5 is below residues involved in canonical ligand binding cavity and it points into the receptor rather than the surface. In addition, prior functional tests for C149 found a similar response [6]. The C149W allele may stop protein function and may have produced octanethiol-specific anosmia.

OR5P3. In the screening assay, Vindija Neandertal did not have a significant response but the human reference responded to five of the seven odorants (Extended Data Fig. 8b). Vindija dose responses, which included higher concentrations of the top three responses for humans (courmarin and both enantiomers of carvone), did not exceed control (Fig. 3). The cell surface expression for Vindija indicated that the OR proteins was present at the cell surface, albeit at a slightly lower level than that in humans. Fst between 1000 Genomes groups and extinct linages (15%) is similar to 1000 Genomes Fst (11.4%) and suggests that 5P3 may be influenced by population structure across evolutionary time (Extended Data Fig. 5). Vindija F159L$^{4.58}$ is in the extracellular part of TM4 (Fig. 3), near 4.53, which is involved in mouse OR trafficking [26]. We observed slightly lower trafficking of the Vindija protein. A similar mutation (S155A$^{4.56}$) in human OR1A2 decreases *in vitro* responses to (S)-(-)-citronellal [25]. If OR5P3 F159L$^{4.58}$ is involved in odorant binding, the mutation of this position from phenylalanine to leucine might prevent the π-π stacking interaction between the aromatic residue and coumarin. We conclude that the Vindija protein is not functional and this might be attributable to

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

many potential reasons (a few examples are lack of odorant binding, or fail in activation mechanism, or fail to bind the G protein).

OR10G3. The screening assay revealed significant responses for all seven odors in the Denisovan OR and in the human reference (Extended Data Fig. 8a). Denisovan variants had significantly stronger dose responses to vanillin and ethyl vanillin compared to humans (Table 1; Fig. 3). Fst was higher between us and extinct lineages (19.5%) compared to within 1000 Genomes (2.4%), suggesting that populations are much less differentiated by this gene today than in the past (Extended Data Fig. 5). Neither of the Denisovan variants ($S73G^{2.52}$, $E197K^{5.36}$) were located in conserved amino acid regions (Fig. 3). TM2 is not involved in odorant binding or receptor function, implying that $S73G^{2.52}$ probably did not change the receptor response. TM5 ($E197K^{5.36}$) forms part of the binding cavity but position 5.36 is located at the very limit of ECL2. K at this position is a rare residue in present-day human ORs (3.6%), suggesting a functional, adaptative reason for this change. The location of the variant suggests it may be involved in ligand entry.

OR10J5. There were three significant screening responses to lyral and helional and eugenol (Extended Data Fig. 8a). The Denisovan response to top odor lyral was lower than that of the human reference (Table 1, Fig. 3). Fst between 1000 Genomes and extinct lineages was 25% (1000 Genomes, 6.8%), again suggesting that the strong differentiation among humans in evolutionary history is much reduced today (Extended Data Fig. 5). The $G21R^{Nter}$ variant of 10J5 found in the Denisovan is located at the very

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

end of the N terminal end, just before the start of TM1. The role of this region in OR function is undetermined.

Discussion

Ten novel missense variants were located in 8 genes (out of 30), with five being functionally different from humans (1C1, 2C1, 5P3, 10G3, 10J5), one the same (1A1) and two without identifiable ligands (2B11, 6P1). Given the small percentage of genes with variants altering OR function, members of the genus *Homo* likely shared an olfactory repertoire, with Neandertals and Denisovans smelling the same range of odors we do but having different dose responses to those odors. When OR function was altered by a novel OR, the difference was in sensitivity rather than specificity. Novel Denisovan OR variants (1A1, 2C1, 10G3, 10J5) were twice as responsive as human equivalents to odors humans perceive as spicy, balsamic, and unpleasant (Table 1), but not to odors perceived as floral. Novel Neandertal variants were three times less responsive than human ORs, including reduced responses to odors perceived as green, floral, and spicy (Table 1). There is some correlation in Neandertal skull morphology that suggests their olfactory bulbs were smaller than humans [3], but the link between bulb size and olfactory acuity is unclear [27,28].

The Denisovan ORs strong response to honey may be ecologically significant. Honey is the most energy-dense natural food and is a prized component of living hunter-gatherer diets (except where bees are rare or absent)—even great apes have a ‘honey tooth’ [29]. Energy-dense foods like sugars are sought by larger-brained primates [30] and, based on

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

oral microbiome data, Neandertals and humans share functional adaptations in nutrient metabolism including starch digestion that are not found in our closest ape relatives [31]. The high response to vanilla odors suggests a potential response to sweet things—an odor-taste pairing common in humans [32].

Local ecological adaptive pressures may have acted on ORs in extinct lineages to produce the few novel variants observed, but extinct lineages were less variable in OR genes and proteins compared to 1000 Genomes. While differences in sample sizes might account for some of the striking differences, the reduced variation is probably due to genetic drift effect or conservation. Small effective population sizes cause bottlenecks and small, isolated populations [20] are prone to allele fixation through genetic drift. Purifying selection has been observed in chimpanzee ORs compared to a mix of relaxed and positive selection in human ORs [33] and extinct lineages may also have been subject to this, as evidenced by fewer variants that mostly code for synonymous proteins (compared to protein variation in 1000 Genomes). The mean of Fst values across genes comparing 1000 Genomes to extinct lineaages (11%) is higher than those for 1000 Genomes population comparisons (4%), which suggests that there are structural differences between them and us that reflect both explanations—drift and conservation. Based on our data the last common ancestor shared by *Homo* (humans, Neandertals, Denisova, and others) and *Pan* (chimpanzees, bonobos) had a conserved set of ORs. Humans derived away from the pattern of conservation more recently, with evolutionary pressure toward increased missense variation.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Despite having a shared repertoire with Neandertals and Denisovans, humans are novel in their highly variable OR repertoire which may reflect cultural adaptations following migrations from Africa. Relaxed selection on OR genes for groups no longer engaging in traditional lifestyles is possible [34]. 1000 Genomes groups outside Africa are less variable in most OR genes than those in Africa and OR gene enrichment is observed in African hunter-gatherer groups (Hadza and Pygmies) but not African agricultural (Yoruba) and pastoral (Maasai) groups [35]. Tanzanian Sandawe hunter-gatherers show no OR allelic enrichment, however, which undermines the case for relaxed selection [35]. High allelic diversity and OR generalization on a broad scale may have functional implications, such as increasing the effective size of the olfactory repertoire [5]. Understanding our unique OR allelic diversity is an important challenge.

Our data provide insights into how the dispersal of human lineages outside of Africa (Denisova, Neandertal, ancient human) may have affected olfactory gene repertoire and function. Understanding the evolutionary genetics and functional significance of observed OR allelic variability in and almong human populations and extinct relatives sheds light on the role of olfaction in key aspects of human culture, and perhaps our current success as a global species.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Comparison of OR activity index for human and extinct lineages*

Gene	Odorant	Description	Note	Denisovan	Human
OR1A1	(+)-carvone	mint, herbal, spearmint	green	10.29	10.64
OR1A1	1-heptanol	musty, leaf, violet, herbal, green	green	0.80	0.89
OR1A1	allyl phenyl acetate	honey, fruity, rum	fruity	13.87	13.12
OR1A1	beta-citronellol	citronella, rose, leafy, oily petal	floral	9.03	7.52
OR1A1	benzyl acetate	floral, jasmin, fruity	floral	2.38	2.89
OR1A1	citral	lemon, juicy, lemon peel	fruity	2.66	2.99
OR2C1	1,3-propanthiol	meaty, sulfurous	sulfur	0.97	0.71
OR2C1	2- propanthiol	cooked vegetable, mustard	sulfur	1.02	0.64
OR2C1	nonanthiol	sulfurous	sulfur	1.45	0.00
OR2C1	octanethiol	sulfurous	sulfur	4.65	0.94
OR10G3	ethyl vanillin	vanilla, creamy, caramellic	balsamic	10.28	3.57
OR10G3	eugenol	clove, spicy, woody	spicy	1.90	1.37
OR10G3	g-undecalactone	peach, creamy, coconut	fruity	1.56	0.94
OR10G3	vanillin	vanilla, creamy, chocolate	balsamic	8.98	3.52
OR10J5	helional	fresh green, watery, ozone	green	1.80	1.53
OR10J5	lyral	floral, muguet	floral	4.33	11.71
				Neandertal Chagyrskaya	**Human**
OR1C1	androstadienone	urine, sweating	animal	3.73	4.58
				Neandertal Vindija	**Human**
OR5P3	(-)-carvone	spicy, bready, caraway	spicy	1.92	15.41
OR5P3	(+)-carvone	mint, herbal, spearmint	green	1.09	13.14
OR5P3	1-heptanol	musty, leaf, violet, herbal, green	green	1.22	1.32
OR5P3	1-hexanol	fusel, oily, fruity, green	green	1.14	1.73
OR5P3	acetophenone	pungent, hawthorn, almond, mimosa	floral	1.30	1.80
OR5P3	coumarin	hay, tonka	balsamic	2.00	26.41

*** Color coding is from low (purple) to high (mustard).**

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

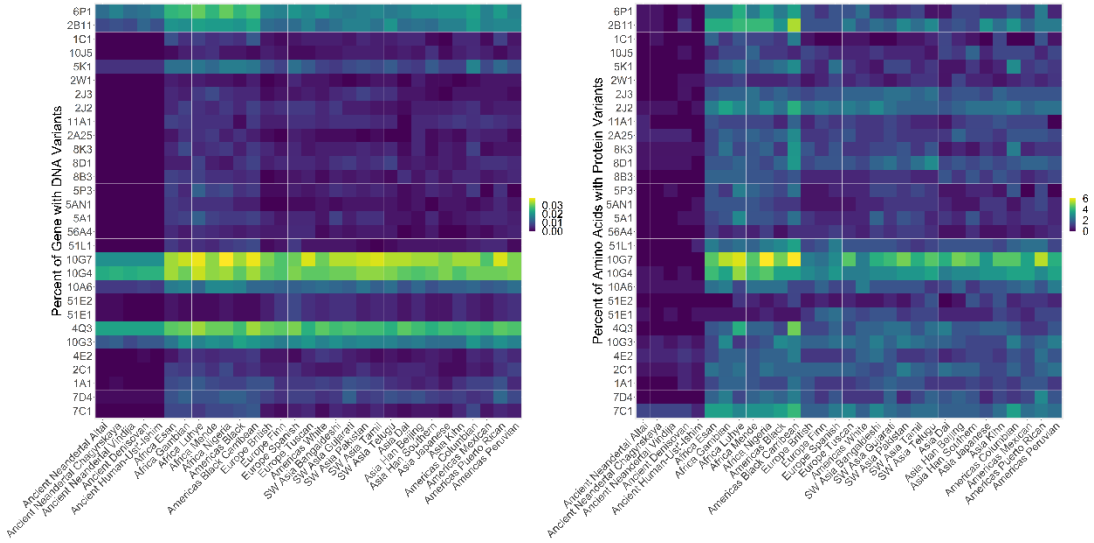


Fig. 1: Percent OR variation (total variant count per gene divided by total basepairs in gene) by population based on raw counts for ancient populations and raw count for the consensus sequence of each gene for each of 26 groups in 1000 Genomes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

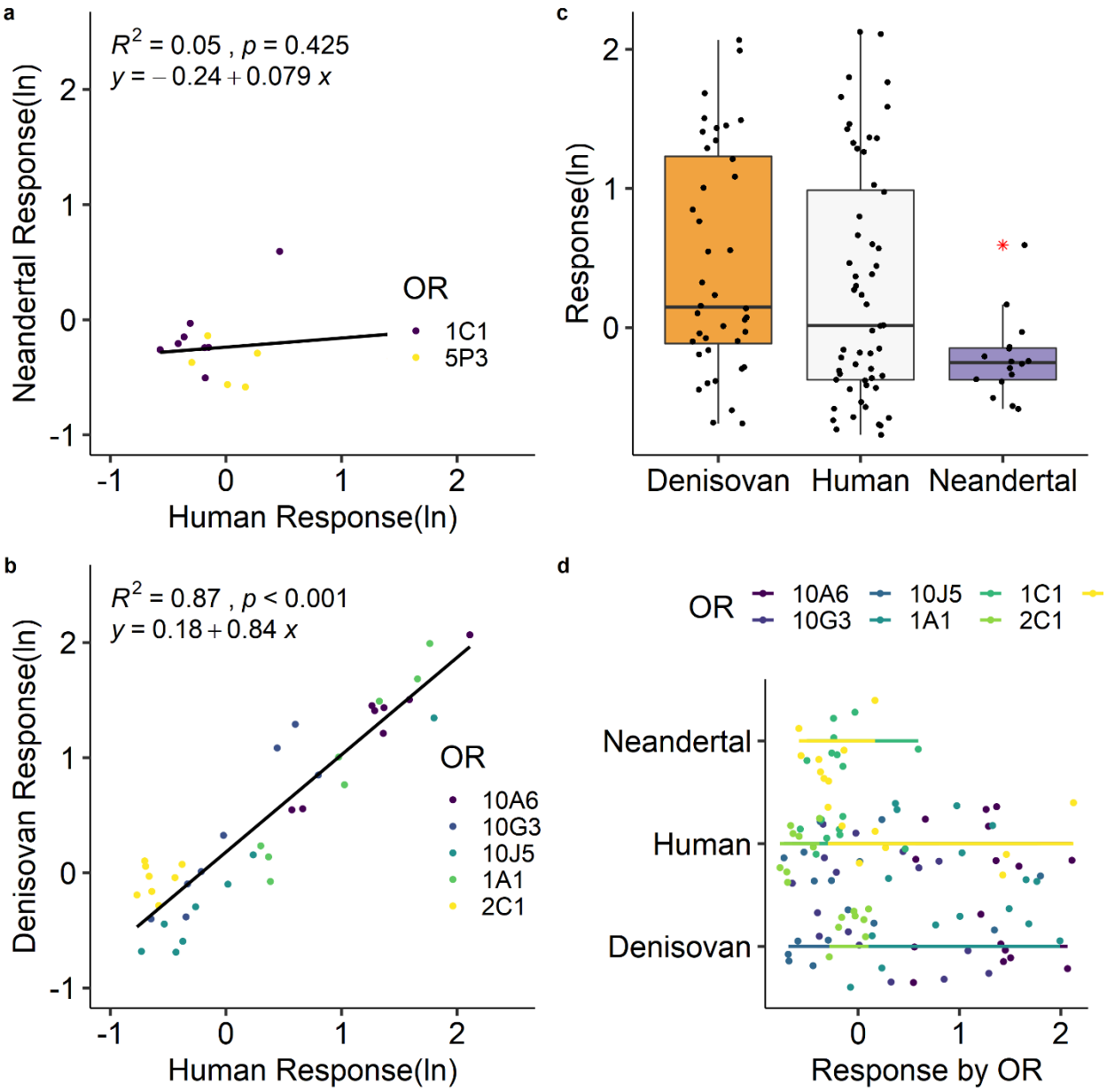


Fig. 2: Regression results for dose responses. a) Human and Denisovan, b) Human and Neandertal, c) Boxplots of dose responses for all samples showing median, box boundaries (first and third quartiles), and two whiskers (upper whisker extends to the largest value no further than 1.5 inter-quartile range from third quartile; lower whisker extends to the smallest value at most 1.5 inter-

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

quartile range of first quartile, outliers identified with red asterisk), d) Dose response (Activity) by OR for all three samples.

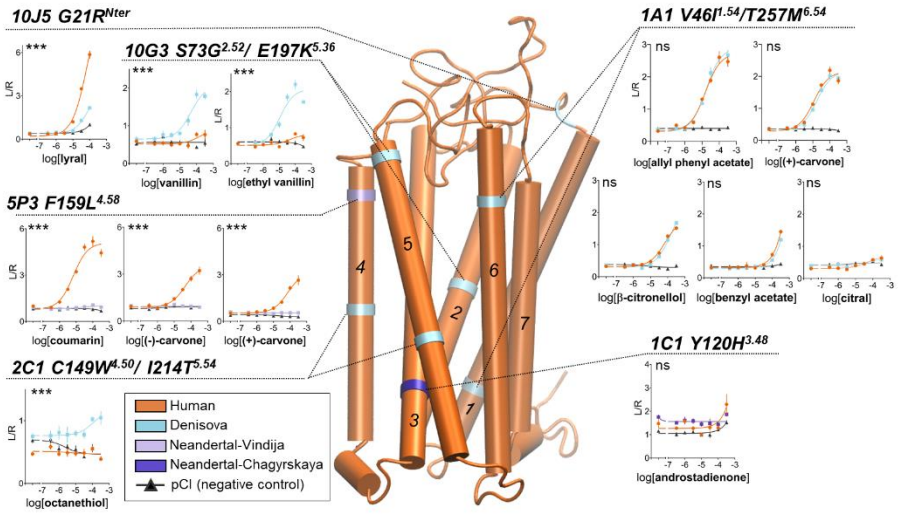


Fig. 3: Homology model of the human consensus odorant receptor to illustrate location of variants. Panels show dose-response for odorants that were significantly activating ORs in screening. (SI Fig. 3; 10G3 and 2C1 both have one shared and one novel mutation). The x-axis of panels represents the odorant concentration (M) and the y-axis the normalized luminescence generated by the activated OR. Error bars are SEM (standard error of the mean).

References

1 Varki, A., Geschwind, D. H. & Eichler, E. E. Human uniqueness: genome interactions with environment, behaviour and culture. *Nature Reviews Genetics* **9**, 749, doi:10.1038/nrg2428 (2008).

2 Winder, I. C. *et al.* Evolution and dispersal of the genus *Homo*: A landscape approach. *J. Hum. Evol.* **87**, 48-65, doi:http://dx.doi.org/10.1016/j.jhevol.2015.07.002 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3 Jacobs, Z. *et al.* Timing of archaic hominin occupation of Denisova Cave in southern Siberia. *Nature* **565**, 594-599, doi:10.1038/s41586-018-0843-2 (2019).

4 Hughes, G. M., Teeling, E. C. & Higgins, D. G. Loss of olfactory receptor function in hominin evolution. *PLoS One* **9**, e84714, doi:10.1371/journal.pone.0084714 (2014).

5 Trimmer, C. *et al.* Genetic variation across the human olfactory receptor repertoire alters odor perception. *Proceedings of the National Academy of Sciences*, 201804106, doi:10.1073/pnas.1804106115 (2019).

6 Mainland, J. D. *et al.* The missense of smell: functional variability in the human odorant receptor repertoire. *Nat. Neurosci.* **17**, 114-120, doi:10.1038/nn.3598 (2014).

7 Hughes, G. M. *et al.* The birth and death of olfactory receptor gene families in mammalian niche adaptation. *Mol. Biol. Evol.*, msy028-msy028, doi:10.1093/molbev/msy028 (2018).

8 Saraiva, L. R. *et al.* A transcriptomic atlas of mammalian olfactory mucosae reveals an evolutionary influence on food odor detection in humans. *Science Advances* **5**, eaax0396, doi:10.1126/sciadv.aax0396 (2019).

9 Luca, F., Perry, G. H. & A., D. R. Evolutionary Adaptations to Dietary Changes. *Annu. Rev. Nutr.* **30**, 291-314, doi:10.1146/annurev-nutr-080508-141048 (2010).

10 McRae, Jeremy F. *et al.* Identification of Regions Associated with Variation in Sensitivity to Food-Related Odors in the Human Genome. *Curr. Biol.* **23**, 1596-1600, doi:http://dx.doi.org/10.1016/j.cub.2013.07.031 (2013).

11 Jaeger, S. R. *et al.* A Mendelian trait for olfactory sensitivity affects odor experience and food selection. *Curr. Biol.* **23**, 1601-1605, doi:10.1016/j.cub.2013.07.030 (2013).

12 McRae, J. F. *et al.* Genetic Variation in the Odorant Receptor OR2J3 Is Associated with the Ability to Detect the “Grassy” Smelling Odor, cis-3-hexen-1-ol. *Chem. Senses* **37**, 585-593, doi:10.1093/chemse/bjs049 (2012).

13 Lunde, K. *et al.* Genetic variation of an odorant receptor OR7D4 and sensory perception of cooked meat containing androstenone. *PLoS One* **7**, e35259 (2012).

14 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

15 Zavala, E. I. *et al.* Pleistocene sediment DNA reveals hominin and faunal turnovers at Denisova Cave. *Nature*, doi:10.1038/s41586-021-03675-0 (2021).

16 Chen, F. *et al.* A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau. *Nature* **569**, 409-412, doi:10.1038/s41586-019-1139-x (2019).

17 Jacobs, G. S. *et al.* Multiple Deeply Divergent Denisovan Ancestries in Papuans. *Cell* **177**, 1010-1021.e1032, doi:10.1016/j.cell.2019.02.035 (2019).

18 Hoover, K. C. *et al.* Global survey of variation in a human olfactory receptor gene reveals signatures of non-neutral evolution. *Chem. Senses* **40**, 481-488, doi:10.1093/chemse/bjv030 (2015).

19 Zhuang, H., Chien, M. & Matsunami, H. Dynamic functional evolution of an odorant receptor for sex-steroid-derived odors in primates. *Proceedings of the National Academy of Sciences* **106**, 21247 - 21251 (2009).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

20 Mafessoni, F. *et al.* A high-coverage Neandertal genome from Chagyrskaya Cave. *Proceedings of the National Academy of Sciences* **117**, 15132-15136, doi:10.1073/pnas.2004944117 (2020).

21 Wright, S. The interpretation of population structure by f-statistics with special regard to systems of mating. *Evolution* **19**, 395-420, doi:https://doi.org/10.1111/j.1558-5646.1965.tb01731.x (1965).

22 Templeton, A. R. Human Races: A Genetic and Evolutionary Perspective. *American Anthropologist* **100**, 632-650 (1998).

23 Saito, H., Chi, Q., Zhuang, H., Matsunami, H. & Mainland, J. Odor coding by a Mammalian receptor repertoire. *Science Signaling* **2**, ra9 (2009).

24 Adipietro, K. A., Mainland, J. D. & Matsunami, H. Functional evolution of mammalian odorant receptors. *PLoS Genet.* **8**, e1002821, doi:10.1371/journal.pgen.1002821 (2012).

25 Schmiedeberg, K. *et al.* Structural determinants of odorant recognition by the human olfactory receptors OR1A1 and OR1A2. *J. Struct. Biol.* **159**, 400-412, doi:10.1016/j.jsb.2007.04.013 (2007).

26 Ikegami, K. *et al.* Structural instability and divergence from conserved residues underlie intracellular retention of mammalian odorant receptors. *Proceedings of the National Academy of Sciences* **117**, 2957-2967, doi:10.1073/pnas.1915520117 (2020).

27 Bastir, M. *et al.* Evolution of the base of the brain in highly encephalized human species. *Nature Communications* **2**, 588, doi:10.1038/ncomms1593 http://www.nature.com/articles/ncomms1593#supplementary-information (2011).

28 Weiss, T. *et al.* Human Olfaction without Apparent Olfactory Bulbs. *Neuron*, doi:10.1016/j.neuron.2019.10.006 (2019).

29 Marlowe, F. W. *et al.* Honey, Hadza, hunter-gatherers, and human evolution. *J. Hum. Evol.* **71**, 119-128, doi:https://doi.org/10.1016/j.jhevol.2014.03.006 (2014).

30 Dominy, N. J. Ferment in the family tree. *Proceedings of the National Academy of Sciences* **112**, 308-309, doi:10.1073/pnas.1421566112 (2015).

31 Fellows Yates, J. A. *et al.* The evolution and changing ecology of the African hominid oral microbiome. *Proceedings of the National Academy of Sciences* **118**, e2021655118, doi:10.1073/pnas.2021655118 (2021).

32 Wang, G., Hayes, J. E., Ziegler, G. R., Roberts, R. F. & Hopfer, H. Dose-Response Relationships for Vanilla Flavor and Sucrose in Skim Milk: Evidence of Synergy. *Beverages* **4**, 73 (2018).

33 Gilad, Y., Man, O. & Glusman, G. A comparison of the human and chimpanzee olfactory receptor gene repertoires. *Genome Res.* **15**, 224-230, doi:10.1101/gr.2846405 (2005).

34 Somel, M. *et al.* A scan for human-specific relaxation of negative selection reveals unexpected polymorphism in proteasome genes. *Mol. Biol. Evol.* **30**, 1808-1815, doi:10.1093/molbev/mst098 (2013).

35 Lachance, J. *et al.* Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. *Cell* **150**, 457-469, doi:10.1016/j.cell.2012.07.009 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

398

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Samples. We cataloged variants in high quality paleogenomic sequence data produced by the Max Planck Institute Leipzig (Extended Data Fig. 2) for Neandertals (Altai, Chagyrskaya, Mezmaiskaya), Denisovan 3 (the only high quality Denisovan genome, finger phalanx), and an ancient human hunter-gatherer from Siberia (Ust'-Ishim). We used the data generated by snpAD in /neandertal/vindija/vcf, an ancient DNA damage-aware genotyper.1 The VCF reference genome for genomes analyzed was hg19/GRCh37. Only variants with a minimum genotype quality of 20 (GC20) were used for downstream analysis. Our sample of living humans was from the 1000 Genomes dataset [36] consisting of over 2,500 individuals from 26 populations in Africa, the Americas, Europe, and Asia (Extended Data Fig. 3).

Although the draft genomes of Neandertals and Denisovans were reported as having more nonfunctional ORs than humans [4], these genomes have since been revised [20,37]. The bioinformatic identification of olfactory pseudogenes has been experimentally challenged [38] by data showing the receptors containing coding sequence regions split into separate exons are conserved across mammals and expressed at the same level as protein-coding receptors with a single exon [39]

Target Genes. We focused on 30 OR genes shown to generate functional response data in previous studies [19] (Supplemental Information Table 1). Gene regions were targeted using NCBI and RefSeq ranges—in the case of eight genes that contained up- and downstream sequences, the region was cut to the protein coding portion of the

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

gene to be consistent with the other 22 genes that contained no up- or downstream areas.

Variant calling. Max Planck and 1000 Genomes Project both used the reference genome, hg19/GRCh37, to call variants using VCFtools [40]. GRCh37 is is built from sequences from different individuals and serves as the wild type relative. If the reference allele is C at a specific genomic position, a variant is called if it is not a C. We excluded insertions and deletions from analysis. BCFtools [41] was used to slice VCFs into chromosomes (ancient DNA) and genes of interest. For 1000 Genomes, BCF tools was also used to generate population specific VCF files. BCF tools was used to create consensus sequences for each gene for the entire 1000 Genomes dataset and for each populations.

We cataloged variants using two sets of VCF data published by Max Planck Institute for Evolutionary Anthropology Leipzig in 2013 and 2016. Both datasets are based on the same original sequence data but the 2016 VCFs were generated using snpAD, an ancient DNA damage-aware genotyper, and are more conservative estimates mutations [37]. The 2016 VCFs include high quality data from three Neandertals (Altai, Chagyrskaya, Vindija), lower quality data from one Neandertal (Mezmaiskaya), high quality from one Denisova (Denisovan 3), and one ancient human contemporary to Altai Neandertal and Denisova in Denisova Cave (Ust'-Ishim). Data are at http://cdna.eva.mpg.de/neandertal/Vindija/VCF. The 2013 VCF data analysis is found in Supplemental Information and included high quality data—but not subject to the

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

damage-aware genotyper—for Altai Neandertal, Denisova 3, and Ust'-Ishim. Data are found at http://cdna.eva.mpg.de/ust-ishim/VCF and http://cdna.eva.mpg.de/neandertal/altai/AltaiNeandertal/.

The 26 groups in 1000 Genomes were comprised of individuals with at least 3 out of 4 grandparents identifying membership in the group. While there are some groups engaging in pastoral or traditional horticulture/farming lifestyles, most practice mixed subsistence economies and lead western lifestyles.

Novel variant calling. R v4.1.0 [42] via R Studio v1.4.1717 [43] were used for analysis. Excel and csv files were read using readxl [44]. The hash function in R [42,43] used 1000 Genomes variant data [36] as a key to flag variants present in ancient samples but not found in present-day humans. Data were manipulated using dplyr [45] and tidyverse [46].

Genetic Analysis. Variants were called using DNAsp to generate Nexus files for each gene comparing a consensus sequence for each of the 26 groups in 1000 Genomes to the human reference sequence used for the published datasets (Max Planck and 1000 Genomes). Consensus sequences were used for each group in 1000 Genomes to reduce the sample size for comparison to ancient sequences which represented individuals, not groups. Biostrings [47] was used in R to create protein sequence files and call amino acid substitutions. Figures for percentage of variation by gene were generated using the R package ggplot2 with the viridis color-blind friendly palette and plotted panels using gridExtra [48]. Concatenated amino acid sequences were used to

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

infer phylogenetic relationships across populations for 30 genes [49]. The phylogenetic tree was created using the R packages ape [50], seqinr [51], phylogram [52], gdsfmt [53], SNPrelate [53], and dendextend [54]. Fst was calculated using gdsfmt [53] and SNPRelate [53].

Primer design. ORs for extinct humans were created by mutating human ORs to match paleogenomic sequence data using chimeric PCR. Forward and reverse PCR primers containing the desired mutation were designed to have a 56ºC or 58ºC annealing temperature, obtained from Integrated DNA Technologies, and diluted to 5μM.

***Chimeric PCR*.** Chimeric PCR was performed using Phusion polymerase and Rho-tagged OR in a pCI vector as a template, with separate reactions using the forward primers paired with a 3' pCI primer or reverse primers paired with a 5' pCI primer [55]. The reaction was started at 98ºC for 30 seconds, then run for 25 cycles of the following conditions: denaturation at 98ºC for 5 seconds, primer annealing at 55ºC for 15 seconds, and elongation at 72ºC for 30 seconds. The reaction was then held at 72ºC for 5 minutes. The products resulting from the forward and reverse primers were combined for each mutant and diluted 10x with distilled water (Gibco). A second PCR was performed using Phusion polymerase, 5' and 3' pCI primers, and the combined and diluted products for each desired mutant as the template. The same PCR conditions were used. The products were purified using the QIAquick PCR Purification Kit (Qiagen), cut with MluI and NotI restriction enzymes, run on a 1.1% agarose gel with GelRed, and extracted using the QIAquick Gel Extraction Kit (Qiagen). The products were then ligated into Rho-tagged pCI cut with MluI and NotI using T4 ligase (New

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

England Biolabs), and used to transform competent ampicillin-resistant E. Coli. These were plated on LB-ampicillin plates and incubated at 37ºC overnight, then a single colony was grown in 4mL of 2XYT-ampicillin (100mg/mL) medium overnight at 37ºC. The Denville Miniprep Purification Kit was used to lyse the bacteria and purify the plasmid DNA. The concentration of the products was determined using an Eppendorf Biophotometer, and then adjusted to 100ng/µL using TE buffer. The products were sequenced using BigDye Terminator Sequencing Kit (Applied Biosystems), purified using Sephadex (GE Healthcare), and sequenced with 3130 Genetic Analyzer (Applied Biosystems).

***Luciferase assay*.** Functionality was tested by a Luciferase assay as described in Zhuang and Matsunami [19]. Hana3A cells were plated in Minimum Essential Medium (MEM) containing 10% FBS (vol/vol), penicillin-treptomycin and amphotericin B on Poly-D-lysine coated 96-well plates (Corning #3843) and co-transfected using Lipofectamine2000 (Invitrogen) with the prepared Rho-tagged OR or the empty vector pCI (negative control), RTP1S (Zhang and Matsunami, 2007), and the muscarinic receptor M3 (Li and Matsunami, 2011). The cells were transfected with elements from the Dual-Glo Luciferase Assay System (Promega): the cAMP response element (CRE) coupled with a luciferase reporter gene (L); and the constitutively active SV40 promoter region coupled with the Renilla luciferase (RL) reporter. In a period of 18-24 hours after transfection, the medium was replaced with 0 µM, 1 µM, 3.16 µM, 10 µM, 31.6 µM, 100 µM, or 316 µM of odorants (Sigma Aldrich) in CD293 (Gibco) containing glutamine and copper and incubated for 3.5 hours. After the addition of Dual-Glo luciferase substrate

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and buffer (Promega), plates were read using BMG Labtech POLARStar Optima plate reader. Promega Stop-and-Glo buffer and RL substrate were added and plates were read again. The degree of activation was quantified for each well in Microsoft Excel using the formula (L-400)/(RL-400) and reported on the y axis (emitted in lumens). The RL quantification is made due to variation in cell numbers in each well and the 0µM values provide the basal activity value for each OR. 0µM thus served as a comparison to identify OR response values to odorant stimulations —if the odor value on the y-axis exceeded the 0µM value, the OR was taken to have responded to the odor but if the odor value did not exceed the 0µM value, the OR was considered unresponsive. Dose responses and EC50s were determined with Graphpad Prism 9. Activity index have been calculated from dose responses by multiplying the absolute value of the logEC50 to the efficacy. If EC50 could not be determined because the dose response did not reached its plateau, logEC50 were set as the arbiruary value -2. These values were used for Table 1 along with odor descriptors and olfactory notes from Good Scents [56]

Flowcytometry. The cell surface expression of the Rho-tagged ORs was evaluated as described in Ikegami, de March et al. [26]. HEK293 cells were plated in 35 mm plates at 25% confluency and grown overnight at 37°C and 5% CO_2. The cells were then transfected with the Rho-tagged ORs and GFP using the Lipofectamine2000 reagent (ThermoFisher Scientific) in MEM supplemented with 10% FBS. 18 to 24h later, the cells were detached from the plates with cellstripper and resuspended in PBS containing 15 mM NaN3, 2% FBS in ice. The cells were then centrifuged at 4°C and resuspended and incubated for at least 30min in primary antibody (mouse anti Rho4D2,

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Millipore Sigma). The cells were then centrifuged at 4°C and resuspended and incubated for at least 30 min and in the dark in phycoerythrin (PE)-donkey anti-mouse F(ab')$_2$ Fragment antibody (Jackson Immunologicals: 715-116-150). Finally, the cells were stained with 7-Amino-actinomycin D (7AAD, Calbiochem). The cell surface expression was monitored through the PE fluorescence emitted by cells GFP positive, single, spherical, viable and 7AAD negative using a BD FACSCanto II. The results were analyzed with Flowjo. Olfr539 and Olfr541 were added to the experiment plan as positive and negative, respectively, controls of OR cell surface expression.

Homology model. The protocol followed a previously published method [55]. 391 human OR sequences were aligned on Jalview to pre-aligned sequences of 11 GPCRs including bovine rhodopsin (PDB: 1U19), human chemokine receptors CXCR4 (PDB: 3ODU) and CXCR1 (PDB: 2LNL), and human adenosine a2A receptor (PDB: 2YDV) using Jalview [57]. The four experimental GPCR structures (1U19, 3ODU, 2YDV and 2LNL) were used as templates to build the human consensus OR by homology modeling with Modeller. The human consensus amino acid sequence was determined by aligning 391 human OR sequences and by selecting the most conserved amino acid for each position. Five models were obtained and the one fulfilling several constraints (binding cavity sufficiently large, no large folded structure in extra-cellular loops, all TMs folded as α-helices, a small α-helix structure between TM3 and TM4) was retained for further structural analysis. Visualization of the models and picture generation were performed with VMD.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Regression Analysis. Data were analyzed in R for Windows 4.1.0, using the R Studio GUI for Windows 1.4.1717. Data were wrangled using dplyr [45] and tidyverse [46]. The dose response dependent variable was fitted to distributions using fitdistrplus [58] and formed log normal distributions. The natural log was used to transform to normally distributed data for linear regression models. The car package [59] was used for ANOVA. 0μM values, as noted above, provided a control (the baseline threshold of OR activation). Because the number of cells varied across plates, the OR tested for each sample had a different 0μM. As such, removing values below the 0μM for each correlation between humans and ancient DNA sample resulted in some odors missing values for one of the lineages. We compared the results of regression for testing using only active ORs (above the 0μM value) and for the full set of ORs—the results were nearly identical (Fig. 2a-b, Extended Data Fig. 10). Testing the full suite of responses, active or below the threshold, allowed a broader comparison of how these cells operate. Analysis was conducted on the full dataset for both between sample correlations and for mean difference testing. Plots for data visualization and results were created using ggplot2 [60] and ggpubr [61]. Panels were created using cowplot [62] and ggarrange.

References

4 Hughes, G. M., Teeling, E. C. & Higgins, D. G. Loss of olfactory receptor function in hominin evolution. *PLoS One* **9**, e84714, doi:10.1371/journal.pone.0084714 (2014).

19 Zhuang, H., Chien, M. & Matsunami, H. Dynamic functional evolution of an odorant receptor for sex-steroid-derived odors in primates. *Proceedings of the National Academy of Sciences* **106**, 21247 - 21251 (2009).

20 Mafessoni, F. *et al.* A high-coverage Neandertal genome from Chagyrskaya Cave. *Proceedings of the National Academy of Sciences* **117**, 15132-15136, doi:10.1073/pnas.2004944117 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

26 Ikegami, K. *et al.* Structural instability and divergence from conserved residues underlie intracellular retention of mammalian odorant receptors. *Proceedings of the National Academy of Sciences* **117**, 2957-2967, doi:10.1073/pnas.1915520117 (2020).

36 The Genomes Project Consortium. The 1000 Genomes Project Consortium. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).

37 Prüfer, K. *et al.* A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*, doi:10.1126/science.aao1887 (2017).

38 Prieto-Godino, L. L. *et al.* Olfactory receptor pseudo-pseudogenes. *Nature* **3**, 93-97, doi:10.1038/nature19824 (2016).

39 Barnes, I. H. A. *et al.* Expert curation of the human and mouse olfactory receptor gene repertoires identifies conserved coding regions split across two exons. *BMC Genomics* **21**, 196, doi:10.1186/s12864-020-6583-3 (2020).

40 Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics (Oxford, England)* **27**, 2156-2158, doi:10.1093/bioinformatics/btr330 (2011).

41 Multiallelic calling model in bcftools (-m), samtools.github.io/bcftools/call-m.pdf.

42 R Development Core Team. Vol. R Foundation for Statistical Computing (Vienna, Austria, 2008).

43 RStudio: Integrated Development for R (RStudio, Inc., Boston, MA, 2015).

44 readxl: Read Excel Files. R package version 1.3.1. https://CRAN.R-project.org/package=readxl (2019).

45 dplyr: A Grammar of Data Manipulation. R package version 0.7.6. https://CRAN.R-project.org/package=dplyr (2018).

46 Wickham, H. *et al.* Welcome to the tidyverse. *Journal of Open Source Software* **4**, 1686, doi:https://doi.org/10.21105/joss.01686 (2019).

47 Efficient manipulation of biological strings. R package version 2.58.0. https://bioconductor.org/packages/Biostrings (2020).

48 gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra (2017).

49 Gadagkar, S. R., Rosenberg, M. S. & Kumar, S. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *J. Exp. Zool. B Mol. Dev. Evol.* **304**, 64-74, doi:10.1002/jez.b.21026 (2005).

50 Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289-290 (2004).

51 Charif, D. & Lobry, J. R. in *Structural approaches to sequence evolution: Molecules, networks, populations Biological and Medical Physics, Biomedical Engineering* (eds U. Bastolla, M. Porto, Roman. H.E, & M. Vendruscolo) 207-232 (Springer Verlag, 2007).

52 Wilkinson, S. P. & Davy, S. K. phylogram: An r Package for Phylogenetic Analysis with Nested Lists. *Journal of Open Source Software* **3**, doi:https://doi.org/10.21105/joss.00790. (2018).

53 Zheng, X. *et al.* A High-Performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data. *Bioinformatics* **28**, 3326-3328, doi:https://doi.org/10.1093/bioinformatics/bts606. (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

54 Galili, T. Dendextend: An r Package for Visualizing, Adjusting, and Comparing Trees of Hierarchical Clustering. *Bioinformatics*, doi:https://doi.org/10.1093/bioinformatics/btv428. (2015).

55 Bushdid, C., de March, C. A., Matsunami, H. & Golebiowski, J. in *Olfactory Receptors: Methods and Protocols* (eds Fabio Marques Simoes de Souza & Gabriela Antunes) 77-93 (Springer New York, 2018).

56 GoodScentsCo. *The GoodScents Company. http://thegoodscentscompany.com/*.

57 de March, C. A., Ryu, S., Sicard, G., Moon, C. & Golebiowski, J. Structure–odour relationships reviewed in the postgenomic era. *Flavour and Fragrance Journal* **30**, 342-361, doi:10.1002/ffj.3249 (2015).

58 Delignette-Muller, M. L. & Dutang, C. fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software* **64**, 1-34 (2015).

59 Fox, J. & Weisberg, S. *An {R} Companion to Applied Regression*. (Sage, 2011).

60 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, 2016).

61 ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.2. https://CRAN.R-project.org/package=ggpubr (2018).

62 cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 1.1.0. https://CRAN.R-project.org/package=cowplot (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements. The high-performance computing and data storage resources operated by the Research Computing Systems Group at the University of Alaska Fairbanks, Geophysical Institute. US National Science Foundation Award #1550409 (KCH), NIH K99DC018333 (CADM), NSF# 1556207 (HM), NIH DC014423 and NIH DC016224

Author contributions.

- KCH contributed to the conception or design of the work, acquisition and analysis of genetic and protein sequence data, interpretation of genetic and functional analysis data, drafted the work and substantively revised the work.
- CADM contributed to the acquisition, analysis, and interpretation of in vitro data including the production of ancient and contemporary OR DNA, protein activation and expression data and protein models, and drafted and substantively revised the work.
- HM contributed to the design and revision of the work
- MC contributed to the design and substantive revision of the work.
- MA contributed to the production of ancient and contemporary OR DNA and acquisition of protein activation data to work.

CADM and KHC contributed equally to the work

Competing interests. The authors declare no competing interests.

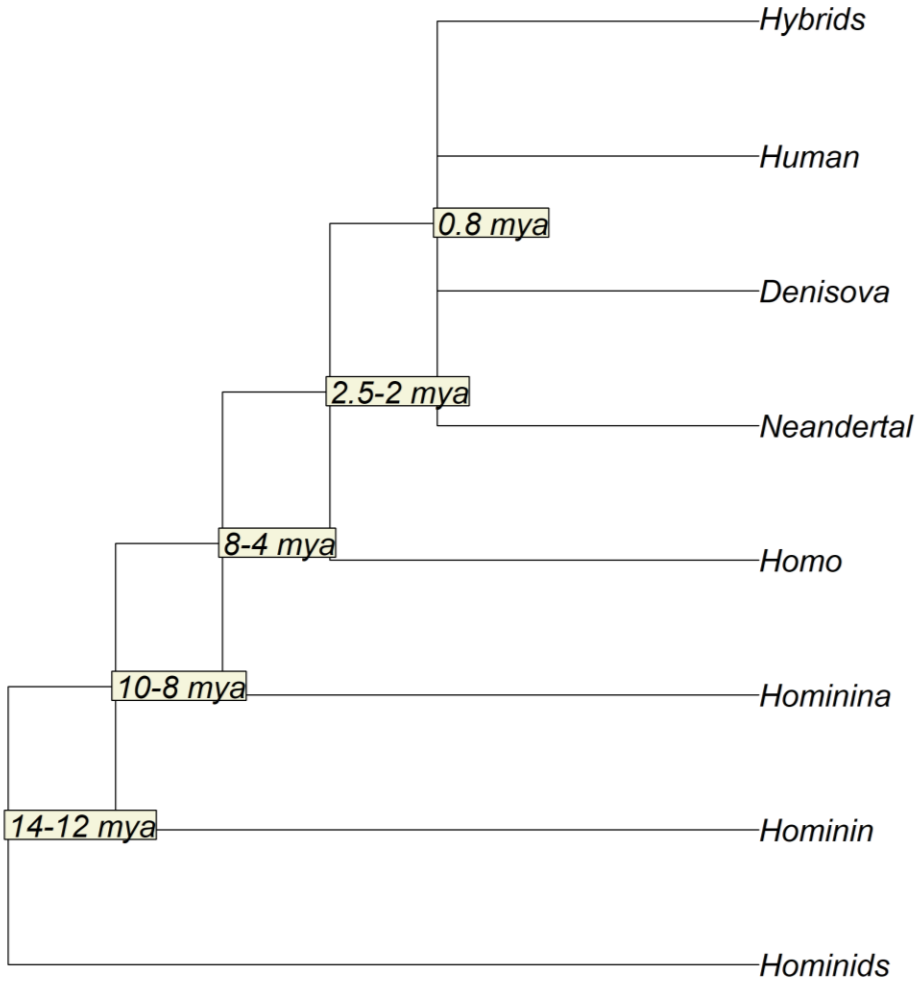
Supplementary Information. Supplementary Information is available for this paper.

Materials & Correspondence. Correspondence and requests for materials should be addressed to Professor Kara C. Hoover, University of Alaska Fairbanks, Fairbanks Alaska 99775 (kchoover@alaska.edu) or Dr. Claire de March, Duke University, Durham, NC 27710 (claire.de.march@duke.edu)

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Extended Data Fig. 1: Phylogenetic tree of Hominids



bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Extended Data Fig. 2: Materials and accession numbers for published genomic data

Lineage	Specimen	Location	Date	Accession/DOI
Denisova	Denisova3	Altai Mts, Siberia	74-82K BP	PRJEB3092 10.1126/science.1224344
Neandertal	Altai	Altai Mts, Siberia	50.3KYA ± 2.2	PRJEB1265 10.1038/nature12886
	Chagyrskaya	Altai Mts, Siberia	781-126 KYA	eva.mpg.de/neandertal/Chagyrskaya/VCF/ 10.1016/j.aeae.2013.07.002
	Mezmaiskaya1	Russian Caucasus	60-70 KYA	PRJEB21195 10.1126/science.aao1887
	Vindija 33.19	Croatia	~50 KYA	PRJEB21157 10.1126/science.aao1887
Ancient Human	Ust'-Ishim	Western Siberia	49K BP	PRJEB6622 10.1038/nature13810
Humans	1000 Genomes	global	living	ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/ 20130502/

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Extended Data Fig. 3: 1000 Genomes populations

	Group	n
Africa	Esan, Nigeria	99
	Gambian	113
	Luhye, Kenya	99
	Mende, Sierra Leone	85
	Yoruban, Nigeria	108
East Asia	Chinese-Beijing Han	103
	Chinese-Southern Han	105
	Chinese-Dai	93
	Japanese	104
Southeast Asia	Kinh, Vietnam	99
Southwest Asia	Bangaldeshi	86
	Gujarati, Houston	103
	Punjabi, Pakistan	96
	Tamil (Sri Lanka), UK	102
	Telugu, UK	102
Europe	American White, Utah	99
	British	99
	Finn	99
	Iberian, Spain	107
	Tuscan	107
Americas	American Black	61
	Caribbean Black	91
	Columbian	94
	Mexicans, LA	64
	Peruvian	85
	Puerto Rican	104

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Extended Data Fig. 4: Summary of Genetic and Protein Variation

	Mean		Range		StDev	
Group	**DNA**	**Protein**	**DNA**	**Protein**	**DNA**	**Protein**
Africa Esan	1.01	1.85	3.20	4.49	0.92	1.04
Africa Gambian	1.03	2.02	3.42	5.45	0.94	1.13
Africa Luhye	1.13	2.18	3.63	5.46	1.01	1.27
Africa Mende	1.03	1.90	3.10	4.17	0.92	0.99
Africa Nigeria	1.08	2.02	3.64	5.78	1.01	1.23
Americas Black	0.97	1.80	3.21	4.53	0.93	1.06
Americas Black Carribean	1.10	2.56	3.63	5.78	1.01	1.39
Americas Columbian	0.92	1.72	3.21	4.49	0.91	1.05
Americas Mexican	0.73	1.37	2.99	3.53	0.83	0.84
Americas Peruvian	0.80	1.45	3.10	3.85	0.89	0.93
Americas Puerto Rican	0.93	1.74	3.52	4.90	0.92	0.97
American White	0.75	1.33	2.90	2.25	0.87	0.70
Ancient Denisova	0.40	0.29	2.46	1.25	0.74	0.35
Ancient Human-Ust'-Ishim	0.39	0.17	2.46	0.64	0.74	0.22
Ancient Neandertal Altai	0.38	0.21	2.35	1.25	0.70	0.29
Ancient Neandertal Chagyrskaya	0.40	0.20	2.46	1.25	0.74	0.29
Ancient Neandertal Vindija	0.36	0.17	2.35	1.25	0.70	0.28
Asia Dai	0.68	1.20	3.42	3.85	0.93	0.86
Asia Han Beijing	0.76	1.34	3.22	4.53	0.87	0.92
Asia Han Southern	0.72	1.41	3.31	4.17	0.86	0.82
Asia Japanese	0.68	1.26	3.10	3.21	0.92	0.82
Asia Kihn	0.75	1.43	3.31	4.21	0.87	0.86
Europe British	0.75	1.25	3.10	2.57	0.91	0.67
Europe Finn	0.70	1.24	2.87	2.56	0.84	0.67
Europe Spanish	0.88	1.48	2.99	3.21	0.88	0.80
Europe Tuscan	0.84	1.44	3.53	4.17	0.89	0.91
SW Asia Bangaldeshi	0.83	1.53	3.31	3.85	0.87	0.88
SW Asia Gujarati	0.83	1.54	3.42	4.17	0.89	0.83
SW Asia Pakistan	0.84	1.53	3.53	4.50	0.92	0.96
SW Asia Tamil	0.81	1.45	3.63	3.53	0.92	0.78
SW Asia Telugu	0.80	1.51	3.42	4.17	0.91	0.97

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Extended Data Fig. 5: Fst for 30 OR genes in extinct lineages and 1000 genomes (1KG) data

Gene	Fst: Among 1KG populations	Gene	Fst: 1KG versus Extinct Lineages
OR5P3	0.114	OR7C1	0.445
OR1C1	0.087	OR1C1	0.379
OR10J5	0.068	OR5AN1	0.279
OR8D1	0.063	OR10J5	0.251
OR2B11	0.061	OR51E2	0.244
OR51E1	0.059	OR51E1	0.197
OR56A4	0.058	OR10G3	0.195
OR4E2	0.056	OR8D1	0.185
OR2A25	0.051	OR5A1	0.185
OR2J3	0.043	OR5P3	0.151
OR51E2	0.042	OR2C1	0.133
OR5A1	0.040	OR10A6	0.133
OR7D4	0.037	OR4E2	0.126
OR1A1	0.034	OR2B11	0.091
OR5K1	0.031	OR10G4	0.074
OR2W1	0.029	OR1A1	0.070
OR2J2	0.028	OR8K3	0.046
OR4Q3	0.028	OR56A4	0.036
OR7C1	0.028	OR2J3	0.013
OR10G7	0.028	OR2A25	0.012
OR51L1	0.025	OR2W1	0.008
OR6P1	0.024	OR51L1	0.000
OR8K3	0.024	OR11A1	0.000
OR10G3	0.024	OR2J2	0.000
OR5AN1	0.023	OR10G7	0.000
OR10G4	0.022	OR7D4	0.000
OR11A1	0.022	OR6P1	0.000
OR2C1	0.015	OR4Q3	0.000
OR8B3	0.010	OR5K1	0.000
OR10A6	0.009	OR8B3	0.000
Mean	**0.0394**	**Mean**	0.108

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

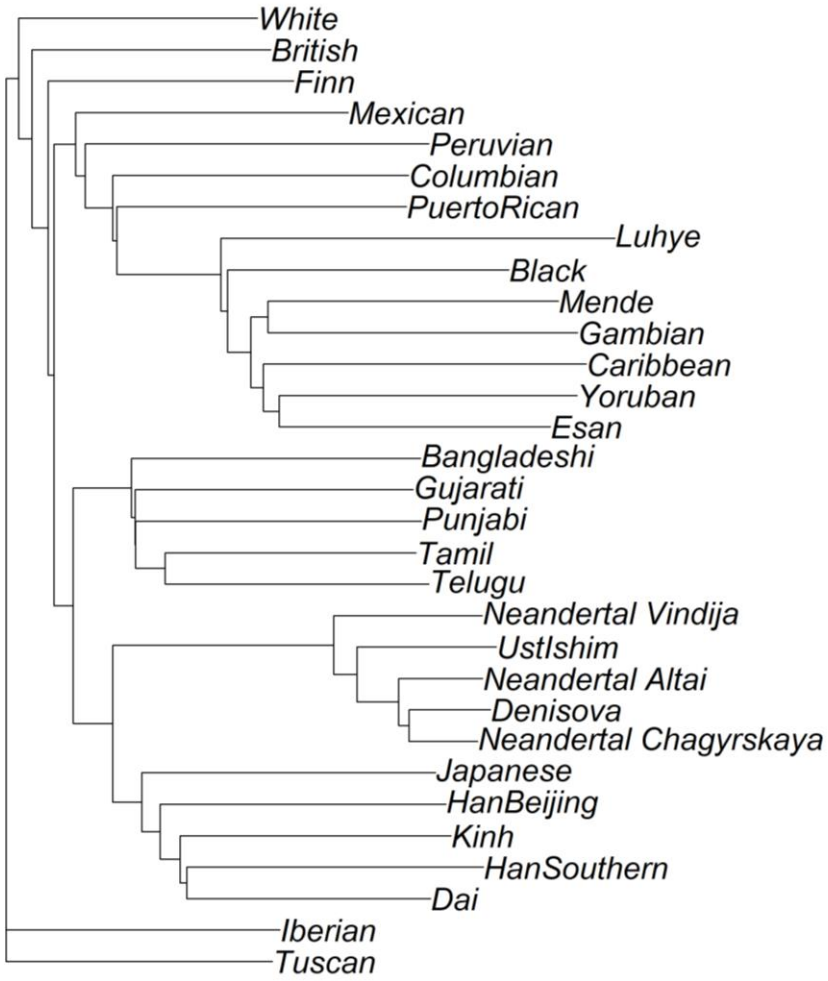
Extended Data Fig. 6: 2016 VCF Novel Variants

Gene	Position	Substitution	Sample	Qual	Depth	GQ	Note
OR1A1	17:3119050	V46I	Denisova	142	34	99	1
	17:3119684	T257M	Denisova	141	37	99	1
OR1C1	1:247921351	Y120H	Chagyrskaya	74	23	74	
OR2B11	1:247614839	V149E	Denisova	148	38	99	1
OR2C1	16:3406581	I214T	Denisova	113	27	99	1
OR5P3	11:7847045	F159L	Vindija	87	18	89	
OR6P1	1:158532749	S216A	Altai	174	47	99	1
			Chagyrskaya	85	17	86	
	1:158533360	F12S	Denisova	102	23	99	1
OR10G3	14:22038287	E197K	Denisova	120	29	99	1
OR10J5	1:159505737	G21R	Denisova	142	48	99	

Gene	Position	Synonymous	Sample	Qual	Depth	GQ	Note
OR2B11	1:247615024	V87	Denisova	541	42	99	1
OR5AN1	11:59132645	A238A	Chagyrskaya	22	17	22	
OR10A6	11:7950075	I45I	Denisova	79	26	79	1
OR10G7	11:123908878	T277	Altai	415	51	99	1
			Chagyrskaya	128	20	99	
			Vindija	96	19	98	

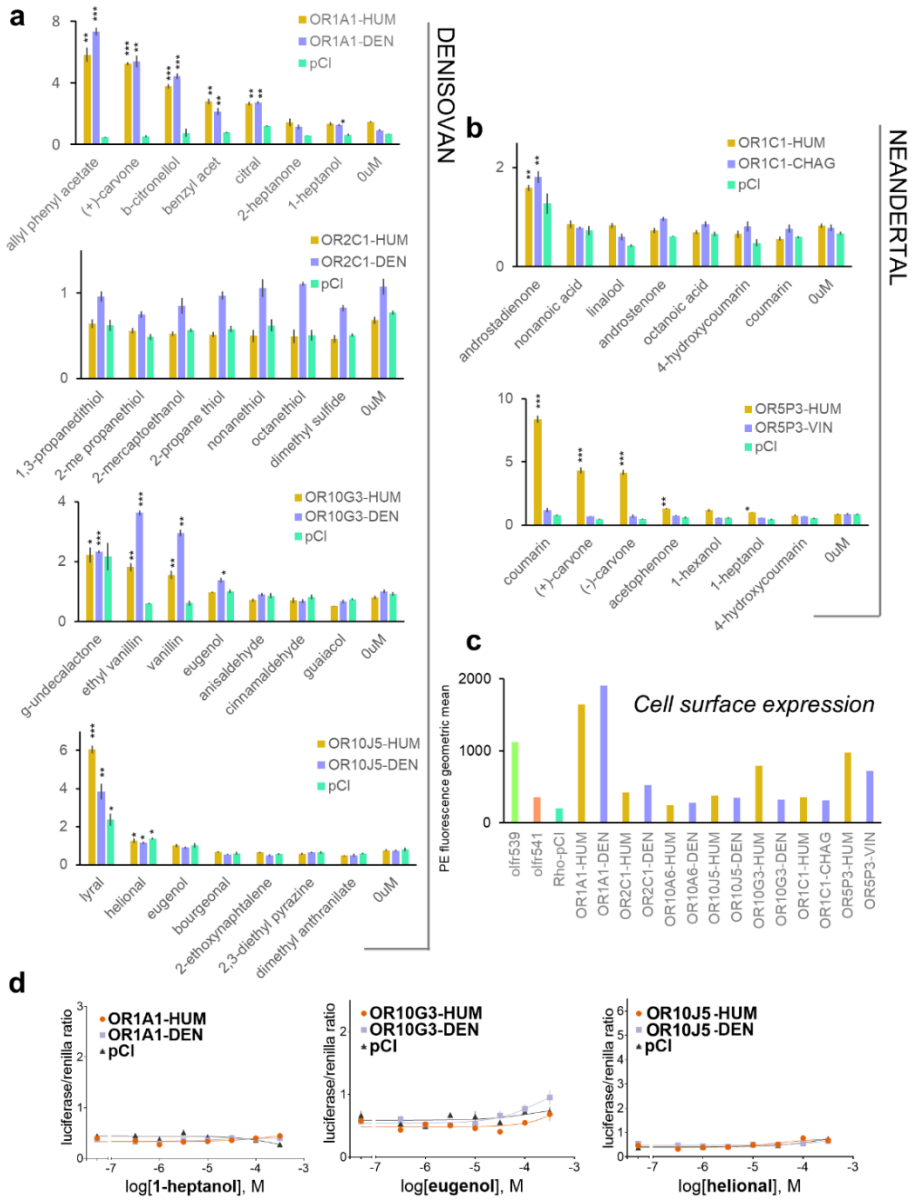
bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Extended Data Fig. 7: Cladogram based on full amino acid sequences for all 30 odorant receptors in ancient lineages and in 1000 Genomes populations.



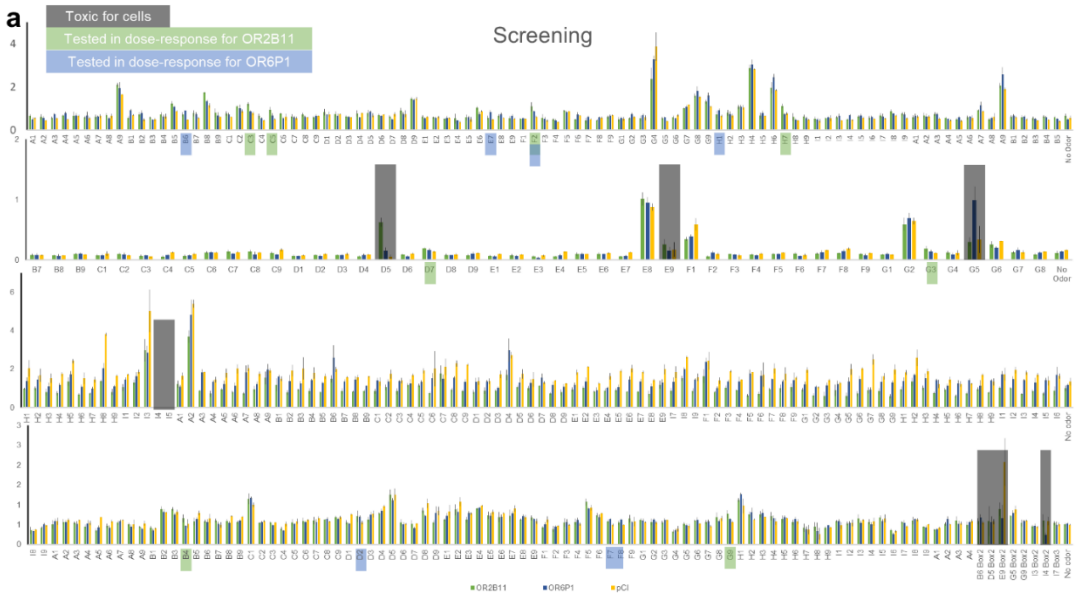
bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Extended Data Fig. 8: Functionality of hominin OR variants. Each Denisova (**a**) and Neandertal (**b**) OR response was first tested in luciferase assay against seven odorants (100µM) previously identified as evoking responses in human OR1A1 [1-3], OR1C1 [1], OR2C1 [2], OR10J5 [1,2], OR5P3 [1,2], and OR10G3 [1]. The empty vector pCI was added to each experiment as a control of OR-specific response. The y-axis represents the normalized luminescence as a measure of OR activation. Asterisks represent the level of significance in a paired t-test between the odorant stimulation and the no odor (0µM) control for each OR (* p<0.05; ** p<0.01; *** p<0.001). **c**. Cell surface expression of all OR variants. Olfr539 and Olfr541, served as positive and negative control of cell surface presence, respectively. **d.** Dose response of the false positives detected in the screening (**a**).



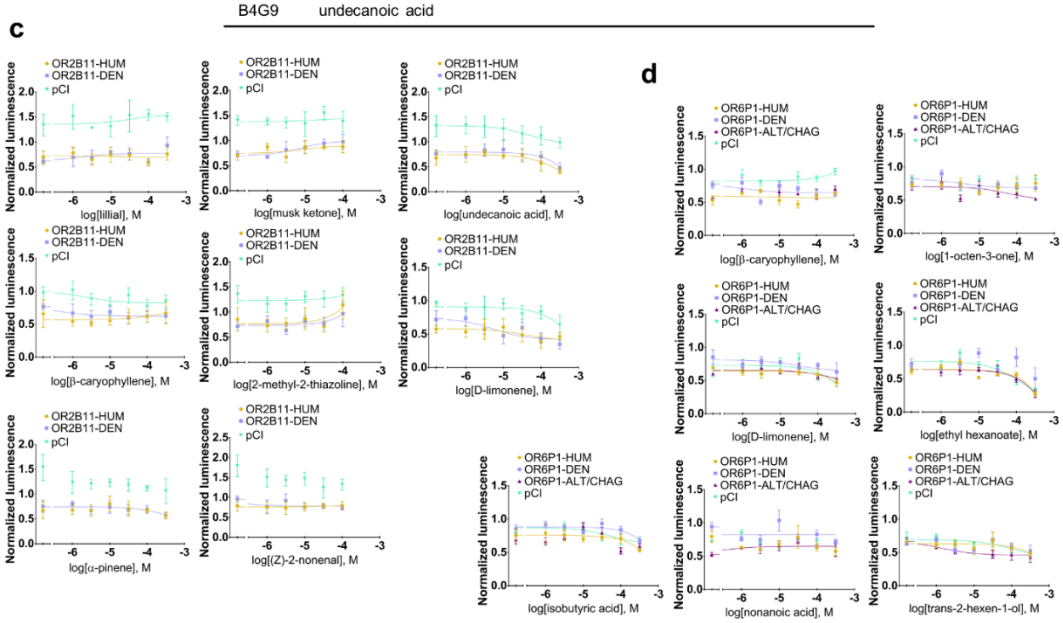
bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Extended Data Fig. 9: OR2B11 and OR6P1 response. a. Activation screening for OR2B11 and OR6P1 against more than 300 odorant molecules at 100 µM. **b**. List of odorants selected for validation by dose response. **c**. Dose-responses of OR2B11 to the odorants listed in **b**. **d.** Dose-responses of OR6P1 to the odorants listed in **b**. In panels **a**, **c** and **d** the y-axis represents the normalized luminescence response that is synonymous with OR activation.



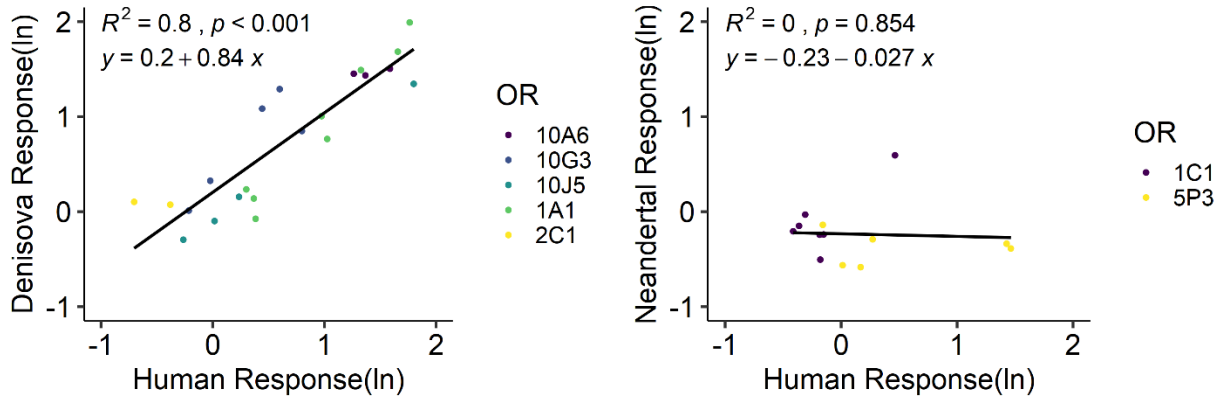
b

OR2B11		OR6P1	
B1H7	Lillal	B1H1	Isobutyric acid
B1F2	Beta caryophyllene	B1F2	Beta caryophyllene
B1C5	Alpha pinene	B1E7	D-Limonene
B1C3	Musk ketone	B1B6	Nonanoic acid
B2D7	2-methyl-2-thiazoline	B4D2	1-Octen-3-one
B2G3	Z-2-nonenal	B4F7	ethyl hexanoate
B4B4	D-Limonene	B4F8	trans-2-hexen-1-ol
B4G9	undecanoic acid		



bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Extended Data Fig. 10: Regression results for active ORs only



bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Information:

Genetic and functional odorant receptor variation in the Homo lineage

Claire A. de March [1,*], Hiroaki Matsunami [1,2] Masashi Abe [1,3], Matthew Cobb [4], Kara C. Hoover [5,*]

[1] Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27710, USA

[2] Department of Neurobiology, Duke Institute for Brain Sciences, Duke University, Durham, NC 27710, USA

[3] Department of Biotechnology and Life Science, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan

[4] Faculty of Life Sciences, The University of Manchester, Oxford Road, Manchester, M13 9PL, USA

[5] Department of Anthropology, University of Alaska Fairbanks, Fairbanks Alaska, 99775, USA

* equal contribution / correspondence

Materials & Correspondence. Correspondence and requests for materials should be addressed to Professor Kara C. Hoover, University of Alaska Fairbanks, Fairbanks Alaska 99775 (kchoover@alaska.edu) or Dr. Claire de March, Duke University, Durham, NC 27710 (claire.de.march@duke.edu)

Contents

Part 1: Supplemental Data for 2016 VCF data (main paper)
Part 2: Functional Expression Analysis of 2013 VCF data

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

PART 1: SUPPLEMENTAL DATA FOR 2016 VCF DATA (MAIN PAPER)

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SI Table 1: Target Genes

Gene	Chr	Region	Odorant	Description
OR6P1	1	158532441-158533394	anisaldehyde	anise, licorice
OR10J5	1	159504868-159505797	lyral	floral, lily
OR2B11	1	247614331-247615284	coumarin	grassy, vanilla, almond
OR1C1	1	247920764-247921708	linalool	floral, bergamot, lavender
OR5K1	3	98188421-98189347	eugenol methyl ether	clove-like
OR2W1	6	29011990-29012952	cis-3-hexen-1-ol	freshly cut grass, leaves
OR2J3	6	29079588-29080603	cinnamaldehyde, cinnamon	cinnamon
OR2J2	6	29141413-29142351	ethyl vanillin	vanilla
OR11A1	6	29394471-29395418	2-ethylfenchol	soil, earth, peat
OR2A25	7	143771313-143772245	geranyl acetate	floral, fruity, rose
OR51E1	11	4673757-4674713	isovaleric acid	sweat, cheese
OR51E2	11	4702979-4703941	androstenone, musk	β-ionone, violet-rose
OR51L1	11	5020213-5021160	allyl phenyl acetate	honey
OR56A4	11	6023281-6024378	undecanal	green coffee, leafy
OR5P3	11	7846584-7847519	coumarin	grassy, vanilla, almond
OR10A6	11	7949265-7950209	3-phenyl propyl propionate	floral balsam hyacinth mimosa
OR8K3	11	56085783-56086721	menthol	mint, menthol
OR5AN1	11	59131932-59132867	muscone	musk
OR5A1	11	59210642-59211589	β-ionone	violet rose
OR10G4	11	123886282-123887217	vanillin	vanilla
OR10G7	11	123908773-123909708	eugenol methyl ether	clove-like (fruits and spices)
OR8D1	11	124179736-124180662	caramel furanone	low-burnt sugar, high-fenugreek/curry
OR8B3	11	124266306-124267247	.+(-) and -(-)carvone	spearmint and carraway

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

OR4Q3	14	20215587-20216528	eugenol methyl ether	clove-like
OR10G3	14	22037934-22038875	vanillin	vanilla
OR4E2	14	22133297-22134238	n-amyl acetate	bananas, apples
OR2C1	16	3405941-3406879	octanethiol	sulphur
OR1A1	17	3118915-3119844	-(-)carvone	spearmint
OR7D4	19	9324575-9325513	androstenone, adrostadienone	musk
OR7C1	19	14909986-14910948	androstadienone	musk

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SI Table 2: 2016 VCF Shared Variants

Gene	Substitution	Position	RS	Specimen	Quality	Depth	GQ	Note
OR1A1	V233M	17:3119611	rs17762735	Ust'-Ishim	503	39	99	1
OR2A25	A209P	7:143771937	rs2961135	Altai	137	35	99	1
				Chagyrskaya	100	22	99	
				Denisova	108	25	99	1
				Vindija	111	26	99	
OR2A25	S75N	7:143771536	rs6951485	Altai	159	40	99	1
				Chagyrskaya	100	20	99	
				Denisova	99	20	99	1
				Vindija	125	29	99	
OR2B11	S12P	1:247615251	rs147030298	Denisova	108	25	99	1
				Altai	213	30	99	1,2
				Chagyrskaya	106	24	99	2
				Denisova	103	23	99	1,2
OR2C1	C149W	16:3406387	rs1218763	Altai	120	29	99	1
				Chagyrskaya	92	31	92	
				Denisova	128	32	99	1
				Ust'-Ishim	144	39	99	1
				Vindija	108	25	99	
OR2J2	A302V	6:29142317	rs41270628	Altai	116	40	99	1
				Chagyrskaya	111	26	99	
				Vindija	81	16	83	
OR2J2	K312M	6:29142347	rs41270632	Ust'-Ishim	109	27	99	1
OR2J3	R226Q	6:29080344	rs3749977	Denisova	125	29	99	1
OR2W1	D296N	6:29012067	rs35771565	Ust'-Ishim	152	41	99	
OR2W1	P138S	6:29012541	rs146659019	Denisova	124	31	99	
OR4E2	Q234R	14:22133997	rs970382	Altai	149	38	99	1
				Chagyrskaya	126	32	99	
				Denisova	147	38	99	1
				Ust'-Ishim	499	39	99	1
				Vindija	109	27	99	
OR4E2	R233H	14:22133994	rs200419591	Denisova	155	39	99	1
OR4E2	V118M	14:22133648	rs2874103	Altai	201	53	99	1
				Chagyrskaya	99	20	99	
				Denisova	163	41	99	1
				Ust'-Ishim	439	47	99	1
				Vindija	131	31	99	
OR5A1	D183N	11:59211188	rs6591536	Ust'-Ishim	109	25	99	1
OR5A1	G173R	11:59211158	rs140829988	Denisova	306	27	99	1
OR5AN1	L289F	11:59132798	rs7941190	Ust'-Ishim	168	47	99	1
OR7C1	E171K	19:14910438	rs10415312	Altai	205	58	99	1
				Chagyrskaya	155	41	99	
				Denisova	165	45	99	1
				Ust'-Ishim	184	52	99	1
				Vindija	140	37	99	
OR7C1	R233K	19:14910251	rs77999564	Altai	235	67	99	1
				Chagyrskaya	150	39	99	

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

				Denisova	175	47	99	1
				Vindija	152	40	99	
OR7C1	S210P	19:14910321	rs16979912	Altai	157	52	99	1
				Chagyrskaya	100	24	99	
				Denisova	147	38	99	1
				Vindija	96	21	98	
OR7C1	V126I	19:14910573	rs10415562	Altai	118	28	99	1
				Chagyrskaya	79	15	80	
				Denisova	87	18	89	1
				Ust'-Ishim	118	30	99	1
				Vindija	64	11	66	
OR8K3	L122R	11:56086147	rs960193	Altai	111	26	99	1
				Chagyrskaya	91	19	93	
				Denisova	87	27	87	1
				Ust'-Ishim	141	38	99	1
				Vindija	108	25	99	
OR10A6	L287P	11:7949350	rs4758258	Altai	52	48	53	1
				Chagyrskaya	133	33	99	
				Denisova	156	41	99	1
				Ust'-Ishim	486	43	99	1
				Vindija	92	21	93	
OR10G3	R236Q	14:22038170	rs536842323	Chagyrskaya	152	23	99	
OR10G3	S73G	14:22038659	rs17792778	Altai	131	32	99	1
				Chagyrskaya	79	15	81	
				Denisova	132	33	99	1
				Ust'-Ishim	334	34	99	1
				Vindija	115	27	99	
OR10G4	A9V	11:123886308	rs11219407	Altai	102	23	99	1
				Chagyrskaya	82	16	84	
				Denisova	82	18	83	1
				Ust'-Ishim	12	4	12	1
				Vindija	40	11	40	
OR10G4	A9V	11:123886307	rs79057843	Altai	224	23	99	1
				Chagyrskaya	102	17	99	
				Denisova	205	20	99	1
OR51E1	S11N	11:4673788	rs17224476	Chagyrskaya	105	22	99	
				Vindija	328	36	99	
OR51E1	V205I	11:4674369	rs185465942	Ust'-Ishim	34	32	34	1
OR51L1	A207V	11:5020832	rs10768450	Ust'-Ishim	270	34	99	1
OR51L1	T196I	11:5020799	rs10768448	Ust'-Ishim	446	45	99	1

Gene	**Substitution**	**Position**	**RS**	**Specimen**	**Quality**	**Depth**	**GQ**	**Note**
OR1A1	Synonymous	17:3119008	rs4325604	Altai	169	45	99	1
				Chagyrskaya	93	20	95	
				Denisova	108	25	99	1
				Ust'-Ishim	169	47	99	1
				Vindija	109	31	99	
OR1C1	Synonymous	1:247921352	rs1552813	Altai	139	35	99	1,2

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

				Chagyrskaya	102	23	99	2
				Denisova	121	29	99	1,2
OR2C1	Synonymous	16:3406516	rs11643487	Altai	124	59	99	1
				Chagyrskaya	115	27	99	1
				Ust'-Ishim	256	35	99	
				Vindija	134	40	99	
OR2W1	Synonymous	6:29012116	rs7341218	Ust'-Ishim	163	43	99	1
OR4E2	Synonymous	14:22133416	rs12717305	Altai	114	37	99	1
				Chagyrskaya	121	29	99	
				Denisova	135	39	99	1
				Ust'-Ishim	665	50	99	1
				Vindija	169	45	99	
OR5A1	Synonymous	11:59211265	rs7941591	Ust'-Ishim	66	28	66	1
OR5P3	Synonymous	11:7847208	rs1482791	Ust'-Ishim	276	31	99	1
OR5P3	Synonymous	11:7847466	rs1482792	Altai	345	51	99	1
				Denisova	168	45	99	1
				Ust'-Ishim	153	51	99	1
OR5P3	Synonymous	11:7847472	rs1482793	Ust'-Ishim	724	53	99	1
OR6P1	Synonymous	1:158533140	rs12080815	Altai	147	38	99	1,2
				Chagyrskaya	91	19	92	2
OR6P1	Synonymous	1:158533224	rs12081915	Altai	175	58	99	1,2
				Chagyrskaya	82	23	82	
OR8D1	Synonymous	11:124179835	rs4936918	Altai	190	52	99	1
				Chagyrskaya	140	35	99	
				Denisova	165	21	99	1
				Vindija	91	19	93	
OR8D1	Synonymous	11:124180082	rs4936919	Altai	126	42	99	1
				Chagyrskaya	133	33	99	
				Denisova	234	22	99	1
				Vindija	104	27	99	
OR10G3	Synonymous	14:22038450	rs11626669	Altai	137	32	99	1
				Chagyrskaya	74	12	75	
				Denisova	102	21	99	1
				Vindija	132	31	99	
OR10G3	Synonymous	14:22038162	rs17197261	Altai	143	45	99	1
				Chagyrskaya	94	20	96	
				Vindija	97	23	97	
OR51E1	Synonymous	11:4674575	rs3817098	Altai	195	52	99	1
				Chagyrskaya	82	14	84	
				Vindija	86	16	87	
OR51E2	Synonymous	11:4703165	rs1123991	Ust'-Ishim	148	38	99	1
OR51L1	Synonymous	11:5020509	rs11035066	Ust'-Ishim	562	48	99	1
OR51L1	Synonymous	11:5020416	rs2445290	Altai	185	54	99	1
				Chagyrskaya	102	27	99	
				Denisova	140	36	99	1
				Ust'-Ishim	223	24	99	1
				Vindija	85	19	86	

[1] also present in 2013 data

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[2] present in 2013 Ust'-Ishim data but not 2016 Ust'-Ishim data

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

PART 2: FUNCTIONAL EXPRESSION ANALYSIS OF 2013 VCF DATA

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We generated sequence data for variant ORs using a custom bioinformatics pipeline (SI Table 3). Original BAM files from the Max Planck Institute for Evolutionary Anthropology (accessible via ftp at http://cdna.eva.mpg.de/neandertal/altai/) were downloaded to the Arctic Region Supercomputing Center. Files were aligned to NCBI Human Genome Build GRCh37/hg19 and split into chromosomes using bamutil.

Altai Neandertal BAM files were organized by chromosome. For Denisova and Ust'-Ishim, samtools view command was used to split BAM files into chromosomes. BAM files for all three genomes were sorted and indexed. The samtools mpileup command was used to call variants and bcftools generated VCF. The VCF file was zipped and indexed using bgzip and tabix (part of the htslib and samtools). Fastas were created using the consensus function.

SI Table 3: Custom variant calling pipeline for data not in VCF format

```
samtools view -b <filename_in>.bam <chromosome number> <filename_out>.bam
samtools sort -l 9 -O bam <filename_in>.bam -o <filename_out>.bam
samtools index <filename_sorted>.bam
samtools mpileup -ufI <filename_RSfasta>.fa <filename_sorted>.bam > <filename>.bcf
bcftools view -vcg <filename_in>.bcf > <filename_out>.vcf
bgzip <filename>.vcf
tabix <filename>.vcf.gz
vcftools --gzvcf <filename>.vcf.gz --chr <#> --from-bp <#> --to-bp <#> --recode --recode-INFO-all --out
<filename>
bgzip <filename>.recode.vcf
tabix <filename>.recode.vcf.gz
cat <filename_RSfasta>.fa | vcf-consensus <filename_in>.recode.vcf.gz > <filename_out>.fa
```

Three genes matched the human wild type (contained no variants, novel or shared) in all three samples tested: OR5K1, OR11A1, OR56A4)—OR11A1 did contain variants but these were outside the coding regions (five in the 3' region and one variant in the 5' region). Six genes were considered as unchanged as compared to contemporary humans: OR10A6, OR10G7, OR1C1, OR51E1, OR51E2, OR51L1. Five genes contained only shared missense mutations: OR2J3, OR5AN1, OR7C1, OR8D1, OR8K3. The remaining 16 genes contained novel missense variants (Supplemental Table9).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SI Table 4: 2013 VCF Variants

In the 2016 VCF data, several variants were also found in the earlier 2013 data. A few variants in the 2013 Ust'-Ishim VCF data were not found in the 2016 dataset—due to the damaged DNA genotyper. Only one variant was not present in the 2013 dataset.

Specimen	Gene	Position	RS	Quality	Raw Read Depth	Mapping Quality	Type
Altai	OR6P1	1:158532749	Novel	228	48	37	S216A
Altai	OR6P1	1:158533140	rs12080815	228	42	39	SYNON
Ust'-Ishim	OR6P1	1:158533140	rs12080815	228	33	36	SYNON
Altai	OR6P1	1:158533224	rs12081915	228	62	37	SYNON
Ust'-Ishim	OR6P1	1:158533224	rs12081915	228	41	37	SYNON
Denisova	OR6P1	1:158533360	Novel	228	20	37	F12S
Ust'-Ishim	OR10J5	1:159505297	rs4656837	222	41	37	SYNON
Denisova	OR10J5	1:159505737	Novel	228	44	37	G21R
Altai	OR2B11	1:247614693	rs6695302	161	33	38	V198M
Denisova	OR2B11	1:247614693	rs6695302	228	23	37	V198M
Ust'-Ishim	OR2B11	1:247614693	rs6695302	228	34	37	V198M
Denisova	OR2B11	1:247614839	Novel	228	37	37	V149E
Altai	OR2B11	1:247614896	rs11583410	228	30	37	I130S
Denisova	OR2B11	1:247614896	rs11583410	228	25	37	I130S
Ust'-Ishim	OR2B11	1:247614896	rs11583410	228	20	37	I130S
Denisova	OR2B11	1:247615024	Novel	222	39	37	SYNON
Denisova	OR2B11	1:247615251	rs147030298	228	25	37	S12P
Ust'-Ishim	OR1C1	1:247921100	rs1552812	228	36	37	SYNON
Altai	OR1C1	1:247921352	rs1552813	228	37	38	SYNON
Denisova	OR1C1	1:247921352	rs1552813	228	29	37	SYNON
Ust'-Ishim	OR1C1	1:247921352	rs1552813	228	33	37	SYNON
Ust'-Ishim	OR5K1	3:98188365	rs9861575	182	50	37	W14C
Ust'-Ishim	OR2W1	6:29012067	rs35771565	228	42	38	D296N
Ust'-Ishim	OR2W1	6:29012116	rs7341218	228	44	38	SYNON
Denisova	OR2W1	6:29012541	rs146659019	228	31	37	P138A
Denisova	OR2J3	6:29080344	rs3749977	228	27	37	R226Q
Altai	OR2J3	6:29080349	rs3130764	228	36	36	I228V
Denisova	OR2J3	6:29080349	rs3130764	228	26	35	I228V
Altai	OR2J3	6:29080450	rs3130765	228	44	31	M261I/M
Denisova	OR2J3	6:29080450	rs3130765	228	35	26	M261I/M
Altai	OR2J2	6:29141772	rs3129158	228	50	31	SYNON
Denisova	OR2J2	6:29141772	rs3129158	203	20	28	SYNON
Altai	OR2J2	6:29142317	rs41270628	228	44	37	A302V
Ust'-Ishim	OR2J2	6:29142347	rs41270632	228	31	34	K312M
Altai	OR2A25	7:143771536	rs6951485	228	45	35	S75N
Denisova	OR2A25	7:143771536	rs6951485	228	22	33	S75N
Denisova	OR2A25	7:143771733	Novel	136	32	37	I141V
Altai	OR2A25	7:143771937	rs2961135	228	43	38	A209P
Denisova	OR2A25	7:143771937	rs2961135	228	25	37	A209P
Ust'-Ishim	OR51E1	11:4674369	rs185465942	33	32	37	V205I
Altai	OR51E1	11:4674575	rs3817098	228	56	37	SYNON

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ust'-Ishim	OR51E2	11:4703165	rs1123991	228	39	37	SYNON
Altai	OR51L1	11:5020416	rs2445290	228	60	38	SYNON
Denisova	OR51L1	11:5020416	rs2445290	228	36	37	SYNON
Ust'-Ishim	OR51L1	11:5020416	rs2445290	156	24	37	SYNON
Ust'-Ishim	OR51L1	11:5020509	rs11035066	222	52	36	SYNON
Ust'-Ishim	OR51L1	11:5020799	rs10768448	222	45	36	T196I
Ust'-Ishim	OR51L1	11:5020832	rs10768450	222	37	37	A207V
Altai	OR5P3	11:7846722	Novel	222	46	37	Y266STOP
Altai	OR5P3	11:7846752	Novel	222	52	37	SYNON
Ust'-Ishim	OR5P3	11:7847208	rs1482791	157	32	37	SYNON
Altai	OR5P3	11:7847466	rs1482792	222	55	39	SYNON
Denisova	OR5P3	11:7847466	rs1482792	228	44	37	SYNON
Ust'-Ishim	OR5P3	11:7847466	rs1482792	228	54	36	SYNON
Ust'-Ishim	OR5P3	11:7847472	rs1482793	222	56	36	SYNON
Altai	OR10A6	11:7949350	rs4758258	228	54	39	SYNON
Denisova	OR10A6	11:7949350	rs4758258	228	42	37	SYNON
Ust'-Ishim	OR10A6	11:7949350	rs4758258	222	45	38	SYNON
Ust'-Ishim	OR10A6	11:7950024	rs12272735	183	28	34	SYNON
Denisova	OR10A6	11:7950075	Novel	228	26	37	SYNON
Altai	OR8K3	11:56086147	rs960193	228	29	38	L122R
Denisova	OR8K3	11:56086147	rs960193	228	24	35	L122R
Ust'-Ishim	OR8K3	11:56086147	rs960193	228	41	36	L122R
Ust'-Ishim	OR5AN1	11:59132798	rs7941190	228	50	37	L289F
Denisova	OR5A1	11:59211158	rs140829988	222	29	37	G173R
Ust'-Ishim	OR5A1	11:59211188	rs6591536	228	25	37	D183N
Ust'-Ishim	OR5A1	11:59211265	rs7941591	228	28	37	SYNON
Altai	OR10G4	11:123886307	rs79057843	175	43	28	A9V
Denisova	OR10G4	11:123886307	rs79057843	160	25	32	A9V
Altai	OR10G4	11:123886308	rs11219407	228	43	28	A9V
Denisova	OR10G4	11:123886308	rs11219407	228	25	32	A9V
Ust'-Ishim	OR10G4	11:123886308	rs11219407	89	11	21	A9V
Ust'-Ishim	OR10G4	11:123886320	Novel	88	7	19	T13A
Altai	OR10G4	11:123886681	Novel	193	59	14	M134V
Denisova	OR10G4	11:123886681	Novel	44	33	10	M134V
Altai	OR10G4	11:123886865	rs4084209	228	31	33	V195E
Denisova	OR10G4	11:123886865	rs4084209	228	19	30	V195E
Ust'-Ishim	OR10G4	11:123886865	rs4084209	228	21	28	V195E
Denisova	OR10G4	11:123886948	Novel	50	32	17	S223T
Denisova	OR10G4	11:123886950	Novel	76	32	17	S223T
Altai	OR10G4	11:123886984	rs4936880	221	50	34	R235G
Denisova	OR10G4	11:123886984	rs4936880	228	19	36	R235G
Altai	OR10G4	11:123887164	rs4936881	228	32	25	K295Q
Denisova	OR10G4	11:123887164	rs4936881	205	17	22	K295Q
Altai	OR10G4	11:123887184	rs4936882	228	28	33	SYNON
Denisova	OR10G4	11:123887184	rs4936882	228	23	34	SYNON

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ust'-Ishim	OR10G4	11:123887184	rs4936882	228	33	33	SYNON
Altai	OR10G7	11:123908878	Novel	222	55	37	SYNON
Altai	OR10G7	11:123908977	rs188316662	143	54	15	SYNON
Denisova	OR10G7	11:123909073	rs11219419	29	30	23	SYNON
Ust'-Ishim	OR10G7	11:123909073	rs11219419	228	25	27	SYNON
Altai	OR10G7	11:123909099	rs59358830	222	37	37	L204I
Ust'-Ishim	OR10G7	11:123909172	rs28414940	74	43	8	SYNON
Altai	OR10G7	11:123909302	rs513591	175	41	33	T136S
Denisova	OR10G7	11:123909302	rs513591	150	17	34	T136S
Altai	OR10G7	11:123909311	rs472442	75	32	30	N133S
Denisova	OR10G7	11:123909311	rs472442	126	15	34	N133S
Ust'-Ishim	OR10G7	11:123909342	rs470242	200	12	27	SYNON
Altai	OR10G7	11:123909637	Novel	70	8	34	SYNON
Ust'-Ishim	OR10G7	11:123909650	rs3894199	111	32	17	G20A
Altai	OR10G7	11:123909653	rs4998340	29	17	22	P19Q
Ust'-Ishim	OR10G7	11:123909653	rs4998340	135	31	18	P19Q
Altai	OR10G7	11:123909656	rs3894198	20	21	21	A18L
Denisova	OR10G7	11:123909656	rs3894198	10	7	19	A18L
Ust'-Ishim	OR10G7	11:123909656	rs3894198	128	28	17	A18L
Altai	OR10G7	11:123909671	rs11827843	113	27	23	T13M
Denisova	OR10G7	11:123909671	rs11827843	47	16	20	T13M
Altai	OR10G7	11:123909695	rs3894197	203	38	35	T5S
Denisova	OR10G7	11:123909695	rs3894197	87	15	26	T5S
Ust'-Ishim	OR10G7	11:123909695	rs3894197	228	24	29	T5S
Altai	OR8D1	11:124179835	rs4936918	228	53	37	SYNON
Denisova	OR8D1	11:124179835	rs4936918	126	19	37	SYNON
Altai	OR8D1	11:124180082	rs4936919	228	45	38	SYNON
Denisova	OR8D1	11:124180082	rs4936919	193	20	37	SYNON
Ust'-Ishim	OR8B3	11:124266906	Novel	221	64	26	SYNON
Altai	OR8B3	11:124266912	rs28398895	228	55	33	SYNON
Denisova	OR8B3	11:124266912	rs28398895	228	34	34	SYNON
Ust'-Ishim	OR8B3	11:124266912	rs28398895	220	61	27	SYNON
Altai	OR8B3	11:124267056	rs144478570	38	61	12	SYNON
Denisova	OR8B3	11:124267056	rs144478570	15	63	12	SYNON
Denisova	OR8B3	11:124267189	rs507335	10	9	2	H20R
Altai	OR4Q3	14:20216298	rs12896533 [1]	228	35	29	F238L
Denisova	OR4Q3	14:20216298	rs12896533 [1]	228	24	28	F238L
Ust'-Ishim	OR4Q3	14:20216298	rs12896533 [1]	228	33	27	F238L
Altai	OR10G3	14:22038162	rs17197261	228	51	38	SYNON
Denisova	OR10G3	14:22038287	Novel	228	27	37	SYNON
Altai	OR10G3	14:22038450	rs11626669	228	35	39	SYNON
Denisova	OR10G3	14:22038450	rs11626669	228	17	37	SYNON
Ust'-Ishim	OR10G3	14:22038525	rs11626693	222	50	37	SYNON
Altai	OR10G3	14:22038659	rs17792778	228	37	39	S73G
Denisova	OR10G3	14:22038659	rs17792778	228	29	37	S73G

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ust'-Ishim	OR10G3	14:22038659	rs17792778	222	36	36	S73G
Altai	OR4E2	14:22133416	rs12717305	228	40	38	SYNON
Denisova	OR4E2	14:22133416	rs12717305	228	38	37	SYNON
Ust'-Ishim	OR4E2	14:22133416	rs12717305	222	51	37	SYNON
Altai	OR4E2	14:22133648	rs2874103	228	57	39	V118M
Denisova	OR4E2	14:22133648	rs2874103	228	40	37	V118M
Ust'-Ishim	OR4E2	14:22133648	rs2874103	222	47	37	V118M
Denisova	OR4E2	14:22133994	rs200419591	228	29	37	R233H
Altai	OR4E2	14:22133997	rs970382	228	39	37	Q234R
Denisova	OR4E2	14:22133997	rs970382	228	31	37	Q234R
Ust'-Ishim	OR4E2	14:22133997	rs970382	222	39	37	Q234R
Altai	OR2C1	16:3405994	rs1228348	228	62	38	SYNON
Ust'-Ishim	OR2C1	16:3405994	rs1228348	165	27	37	SYNON
Altai	OR2C1	16:3406387	rs1218763	228	33	37	C149W
Denisova	OR2C1	16:3406387	rs1218763	228	31	37	C149W
Ust'-Ishim	OR2C1	16:3406387	rs1218763	228	41	37	C149W
Altai	OR2C1	16:3406516	rs11643487	228	59	37	SYNON
Ust'-Ishim	OR2C1	16:3406516	rs11643487	146	36	37	SYNON
Denisova	OR2C1	16:3406581	Novel	228	29	37	I214T
Altai	OR1A1	17:3119008	rs4325604	228	46	36	SYNON
Denisova	OR1A1	17:3119008	rs4325604	228	19	37	SYNON
Ust'-Ishim	OR1A1	17:3119008	rs4325604	228	48	37	SYNON
Denisova	OR1A1	17:3119050	Novel	228	30	37	V46I
Denisova	OR1A1	17:3119383	Novel	228	40	37	L157F
Altai	OR1A1	17:3119397	rs769425	228	59	37	SYNON
Ust'-Ishim	OR1A1	17:3119397	rs769425	222	39	35	SYNON
Ust'-Ishim	OR1A1	17:3119611	rs17762735	222	40	37	V233M
Denisova	OR1A1	17:3119684	Novel	228	27	37	T257M
Denisova	OR7D4	19:9324904	rs546470301	221	35	26	A204T
Denisova	OR7D4	19:9324971	Novel	139	36	31	SYNON
Altai	OR7C1	19:14910251	rs77999564	228	66	37	R233K
Denisova	OR7C1	19:14910251	rs77999564	228	43	37	R233K
Altai	OR7C1	19:14910321	rs16979912	228	55	37	S210P
Denisova	OR7C1	19:14910321	rs16979912	228	38	37	S210P
Altai	OR7C1	19:14910438	rs10415312	228	58	37	E171K
Denisova	OR7C1	19:14910438	rs10415312	228	36	37	E171K
Ust'-Ishim	OR7C1	19:14910438	rs10415312	228	52	37	E171K
Altai	OR7C1	19:14910573	rs10415562	228	36	36	V126I
Denisova	OR7C1	19:14910573	rs10415562	228	24	33	V126I
Ust'-Ishim	OR7C1	19:14910573	rs10415562	228	44	31	V126I

[1] rs12896533 is no longer considered valid (AKA rs61086986)

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

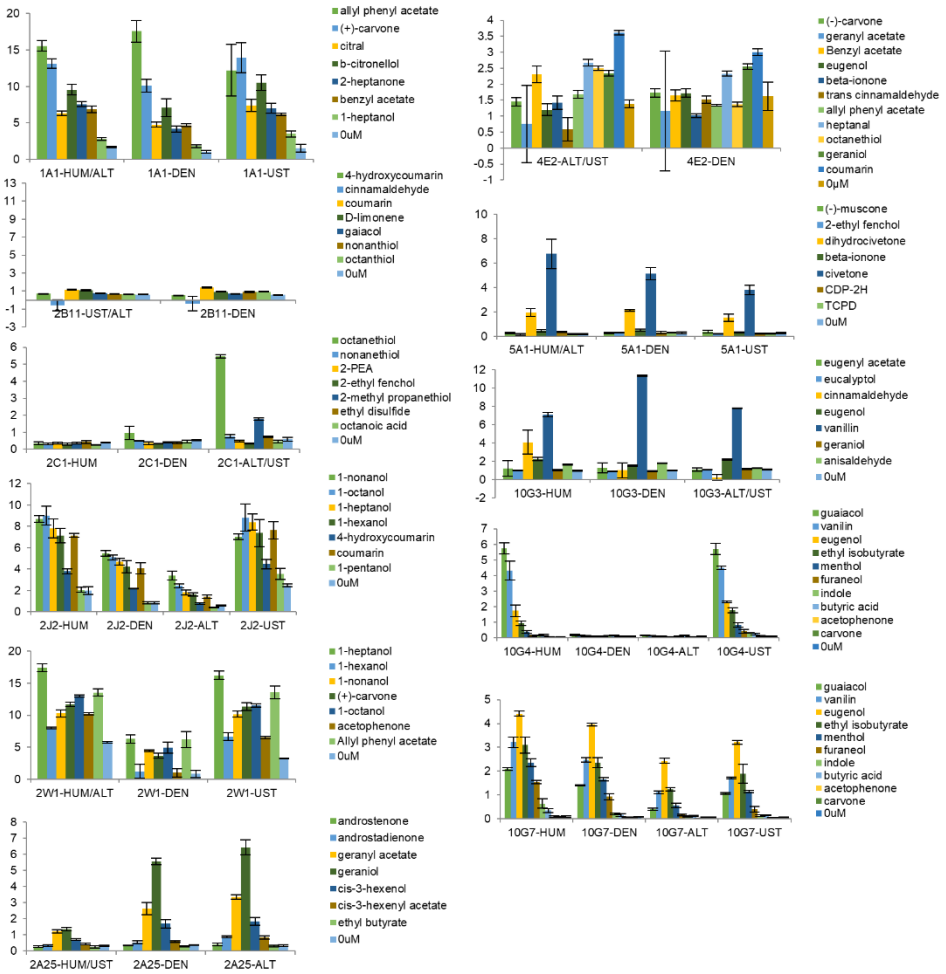
Functional Testing. Functional testing used the same methods presented in the main paper. The total number of genes containing novel variants was 16 (SI Table 5). Three of the 17 genes with novel variants are not reported here because they were either 1) already reported in the main paper (OR6P1, OR10J5) or 2) have been previously published (OR7D4). The novel variant in OR5P3 was not tested because it was a premature stop codon that would have terminated the protein. Several ancient genes shared all variants with modern humans and were not tested. OR10G7 was an exception due to the unique combination of shared variants across ancient samples that were not found in modern humans—while other genes had combinations of shared variants, these were limited to four or less variants, typically only one or two. OR2B11 and OR4E2 contained variants that were not tested because the assay was not responsive (OR2B11) or was minimally responsive (OR4E2). Adding all variants into the assay would not have changed that outcome. A few OR2A25 variants found in contemporary humans were not tested (SI Fig. 1).

SI Table 5: Novel 2013 VCF Variants

Specimen	Gene	Position	RS	Quality	Raw Read Depth	Mapping Quality	Type
Altai	OR10G4	11:123886681	Novel	193	59	14	M134V
Altai	OR10G7	11:123908878	Novel	222	55	37	SYNON
Altai	OR10G7	11:123909637	Novel	70	8	34	SYNON
Altai	OR5P3	11:7846722	Novel	222	46	37	Y266STOP
Altai	OR5P3	11:7846752	Novel	222	52	37	SYNON
Altai	OR6P1	1:158532749	Novel	228	48	37	S216A
Denisova	OR10A6	11:7950075	Novel	228	26	37	SYNON
Denisova	OR10G3	14:22038287	Novel	228	27	37	SYNON
Denisova	OR10G4	11:123886681	Novel	44	33	10	M134V
Denisova	OR10G4	11:123886948	Novel	50	32	17	S223T
Denisova	OR10G4	11:123886950	Novel	76	32	17	S223T
Denisova	OR10J5	1:159505737	Novel	228	44	37	G21R
Denisova	OR1A1	17:3119050	Novel	228	30	37	V46I
Denisova	OR1A1	17:3119383	Novel	228	40	37	L157F
Denisova	OR1A1	17:3119684	Novel	228	27	37	T257M
Denisova	OR2A25	7:143771733	Novel	136	32	37	I141V
Denisova	OR2B11	1:247614839	Novel	228	37	37	V149E
Denisova	OR2B11	1:247615024	Novel	222	39	37	SYNON
Denisova	OR2C1	16:3406581	Novel	228	29	37	I214T
Denisova	OR6P1	1:158533360	Novel	228	20	37	F12S
Denisova	OR7D4	19:9324971	Novel	139	36	31	SYNON
Ust'-Ishim	OR10G4	11:123886320	Novel	88	7	19	T13A
Ust'-Ishim	OR8B3	11:124266906	Novel	221	64	26	SYNON

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

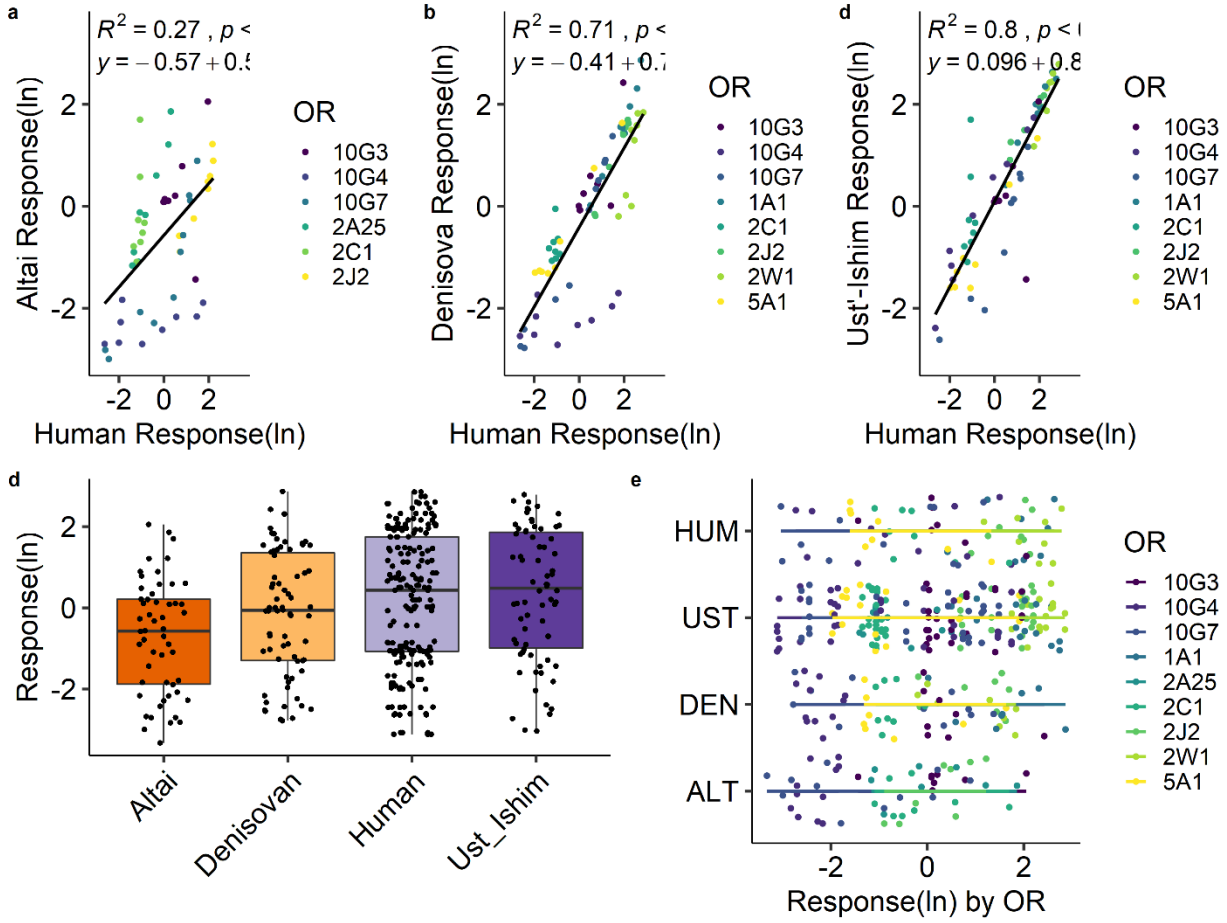
Screening. OR1A1, OR2B11, and OR4E2 responded in a similar way across all samples. Altai Neandertal/Ust'-Ishim OR2C1 responded to octanthiol and 2-methylpropanethiol whereas the modern human and Denisovan version appeared not to be functional. The modern human and Ust'-Ishim versions of OR2J2 responding simillarly to odorants while the corresponding Denisova and Altai Neandertal ORs showed decreased responses. All but Denisova 2W1 responded to the entire set of tested odorants. Denisova 2W1 showed lower responses and, for two of the seven odorants, no response. OR2A25 was more responsive in its Altai Neandertal and Denisova versions than in the modern human/Ust'-Ishim version. OR5A1 responded to dihydrocivetone and civetone for all the considered lineages but was lower in Denisova and Ust'-Ishim than in modern humans/Altai NeandertalAI. OR10G3 responded only to vanillin; the Denisova version showed a higher response than that of modern humans and Altai Neandertal/Ust'-Ishim. For OR10G4, only the modern human and Ust'-Ishim versions were functional; Denisova and Altai Neandertal ORs did not respond to any of the tested compounds. All the versions of OR10G7 were responsive to the tested odorants, although modern humans and Denisova ORs tended to show higher responses than Ust'-Ishim and Altai Neandertal ORs.



SI Fig. 1: Screening of variants identified with the 2013 VCF Data and their corresponding human version response against odorants at 100μM. The y-axis shows normalized luminescence recorded in the luciferase assay.

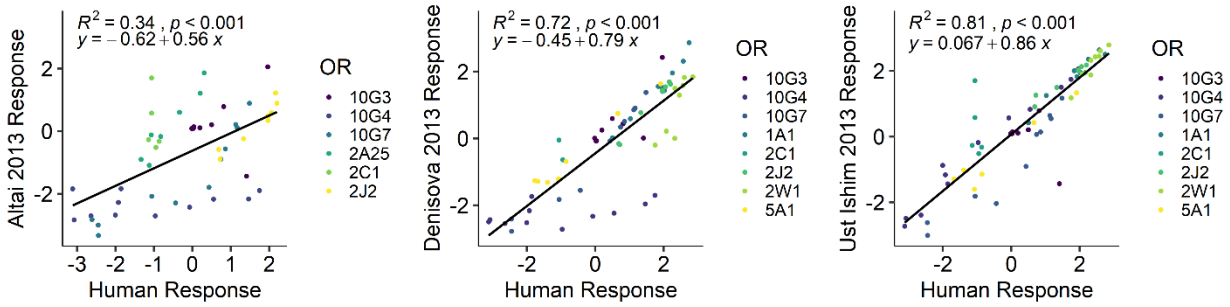
bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Functional Differences. Only data including OR positive response are shown in SI Fig 2. Linear regressions comparing contemporary humans to Altai Neandertal (SI Fig. 2a), Denisova (SI Fig. 2b), and the ancient Siberian human hunter-gatherer Ust'-Ishim (SI Fig. 2c), revealed significant differences. A global comparison to contemporary humans indicated that Altai is significantly different from humans and Denisova (SI Fig. 2d) but there were few differences in terms of patterns of response.



SI Fig. 2: Comparison of human (x-axis) and hominin (y-axis) OR responses **(a)** Altai Neandertal **(b)** Denisova **(c)** Ancient human hunter-gatherer, Ust'-Ishim, **(d)** Boxplots of natural log of OR responses across lineages (humans are divided into ancient and modern). **(e)** OR response by lineage and gene. Each OR is represented by a different color and each point represents the natural log of the response to an odorant. Dotted lines corresponds to the linear regression for the entire set of ORs responses for a given lineage. The corresponding equation and R^2 values are shown on the regression plot.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.13.460146; this version posted September 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



SI Fig. 3. Regression results for active ORs only, 2013 data.

References

1 Adipietro, K. A., Mainland, J. D. & Matsunami, H. Functional evolution of mammalian odorant receptors. *PLoS Genet.* **8**, e1002821, doi:10.1371/journal.pgen.1002821 (2012).

2 Saito, H., Chi, Q., Zhuang, H., Matsunami, H. & Mainland, J. Odor coding by a Mammalian receptor repertoire. *Science Signaling* **2**, ra9 (2009).

3 Schmiedeberg, K. *et al.* Structural determinants of odorant recognition by the human olfactory receptors OR1A1 and OR1A2. *J. Struct. Biol.* **159**, 400-412, doi:10.1016/j.jsb.2007.04.013 (2007).