

# Diversity Within Olfactory Receptor Gene Families: A Case Study of Humans, Mice, Rats, and Elephants



Fethi A. Inan<sup>1</sup>, Ravali Gopaldas<sup>2</sup>, Luis N. Marengo<sup>3</sup>, & Chiquito J. Crasto<sup>2, \*</sup>

<sup>1</sup>Department of Curriculum and Instruction, College of Education, Texas Tech University, Lubbock, Texas

<sup>2</sup>Center for Biotechnology and Genomics, Texas Tech University, Lubbock, Texas

<sup>3</sup>Evolvable Solutions, Sarasota, Florida

**Submission:** December 30, 2023; **Published:** January 10, 2024

**\*Corresponding author:** Chiquito J Crasto, Center for Biotechnology and Genomics, Texas Tech University, Lubbock, Texas, USA

## Abstract

Olfactory receptor (OR) gene families are the largest gene families for all mammals with published genomes. We performed a clustering analysis of the protein sequences of all identified functional olfactory receptor genes for *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, and *Loxodonta africana*. To accomplish this, we used the k-means clustering methodology. The best resolved separation of olfactory receptor genes was achieved by distributing the reduced-dimension OR sequence representations into 18 clusters. Olfactory receptor genes for the four organisms were present in all the 18 clusters, with differing populations for the organisms in each cluster. The analysis did not presuppose gene characteristics such as chromosomal locations or previously identified gene families and subfamilies but focused only on the OR protein sequences. The results show the distribution of clusters in olfactory space with the tabulation of the number of olfactory receptors per cluster per organism. We conclude from the assessments, supported by previous work, of these four organisms with well characterized olfactory repertoires that the overall olfactory space for mammals is representative of all mammals.

**Keywords:** Diversity; Genomes; Olfactory Receptors

## Introduction

The discoverers of olfactory receptors (ORs) in the early 1990s [1] were awarded the Nobel Prize in Medicine or Physiology in 2004 [2]. An added impetus to research in the chemical senses was provided after the publication of the human genome [3]. It was discovered that olfactory receptors constituted the largest gene family in the human genome [4-6]. Subsequent publications of other mammalian olfactory sub-genomes: mouse [7-8], rat, dog [9], and, relatively recently, that of the African bush elephant [10] proved that indeed, the mammalian olfactory genes belong to super-families. From the perspective of the chemical senses, this is likely evolutionarily justified. Olfactory receptors are the first line of interaction with odorants and odors (combinations of odorants) that result in a smell response. A few hundred receptors in each of the mammalian olfactory repertoires are responsible for discriminating several thousand odors. The sense of smell was likely responsible for the fight-or-flight response and the

acquisition of nourishment: both necessary for an organism's survival [11-13].

Interestingly, more than half of the receptors mined from these mammalian genomes have been identified as non-functional or pseudogenic. One likely theory posits that as species (particularly humans) evolved, and other faculties became more finely tuned and developed, species relied less on the sense of smell for survival. [14-16] A lesser-known theory is currently being explored that genetic variants among olfactory receptor genes are distributed based on different racial demographics and different geographic locations [17].

In the current work, we explored the protein sequence variation among functional olfactory receptor genes for *H. sapiens*, *M. musculus*, *R. norvegicus*, and *L. africana*. The protein sequences were clustered based on their sequence similarities and the distribution pattern was visualized using multisequence

alignment tool (Clustal-Omega), t-distributed stochastic neighbor embedding (tSNE) and k-means clustering methods. Distance matrix values between the genes of four different species were calculated by Clustal-Omega. The dimensionality of the data for each gene protein sequence was reduced to two dimensions using tSNE. K-means clustering was used to cluster the data and obtain a graphical representation to understand their distribution in two-dimensional space—namely, an illustration of the olfactory space.

**Methods**

i. The olfactory protein sequences were downloaded from the Olfactory Receptor Database (ORDB-<https://ordb.biotech.ttu.edu/ORDB/>) using the EAV/CR- (Entity Attribute Values with Classes and Relationships) driven database-architecture search system. (<https://ordb.biotech.ttu.edu/ORDB/mam> after [18,19]). This search system identifies information from ORDB based on attributes for a specific class (in this case olfactory receptors) and retrieves the values associated with these attributes. The search for the sequences needed for this analysis was defined thusly CR (Chemosensory Receptor) Type–ORL (Olfactory Receptor Like); Organism–Homo sapiens, Rattus norvegicus, Mus musculus, Loxodonta africana; Type of Sequence–mRNA, cDNA and Genomic DNA; Amino Acids, Nucleotides. This search resulted in the following information: ORLXXX (ORDB identifier), Amino Acid Sequence, Nucleotide Sequence for each OR sequence that met the search filter/criteria.

ii. Nucleotide sequences were included in the search results. For some sequences, the amino acids are not available in ORDB. Each nucleotide sequence was then entered into the search field in the ExpASY [20] Translate tool (<https://web.expasy.org/translate/>). The nucleotide sequences were translated into protein sequences. Any sequences that were identified as pseudogenic (presence of a stop codon in an incomplete sequence) were discarded from further analysis.

iii. Clustal-Omega: The gene sequences were aligned using the Clustal Omega multisequence alignment tool, which is integrated into the EMBL European Bioinformatics Institute Job Dispatcher tools (<https://www.ebi.ac.uk/jdispatcher/msa/>

clustalo) [21]. The platform was used to calculate pairwise similarity values between the genes of four different species. This resulted in a comprehensive distance matrix, with each score ranging between 0 and 1, with a score of 0 indicating 100% sequence identity. Clustal Omega was chosen over other sequence alignment tools because of its ability to handle large numbers of protein sequences, as very large alignment problems can be solved very quickly by using the mBED algorithm [21,22].

iv. t-SNE: t-distributed stochastic neighbor embedding (tSNE), a nonlinear dimensionality reduction method was then applied to the distance matrix determined from the above step [23]. This approach projected the gene sequence similarities onto two-dimensional space while preserving the local relationships and nonlinear structures present in the original similarity data. The Euclidean distance between each gene pair was calculated, and then tSNE was applied to the data to reduce the high-dimensional similarity data into a two-dimensional representation. This allowed the data to be prepared as input into the next clustering and visualizing step.

v. K-means: k-means clustering [24] was used to identify clusters from the two-dimensional tSNE data. The olfactory receptor genes with similar sequence patterns were clustered. The number of clusters is determined by considering the number of members of the olfactory receptor gene family. The resulting 18 clusters were analyzed and interpreted as part of comparing the gene distribution of different species in each cluster to identify potential patterns.

**Results**

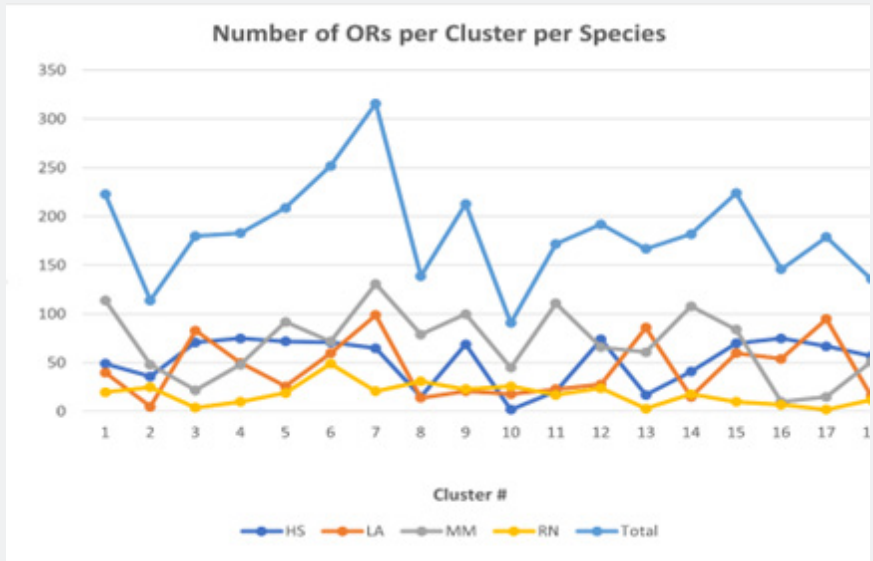
The results of the cluster analysis with 18 clusters are presented in (Table 1) and (Figures 1, 2a&2b). The table and figures illustrate the number of ORs in each cluster and how each species is represented. 947 human ORs, 793 elephant ORs, 1257 mouse ORs, and 321 rat ORs were extracted and downloaded from ORDB. These 3318 ORs were included in the clustering analysis, which followed a multiple sequence analysis using the Clustal-Omega software.

**Table 1:** The table shows the specific number of ORs in each cluster for

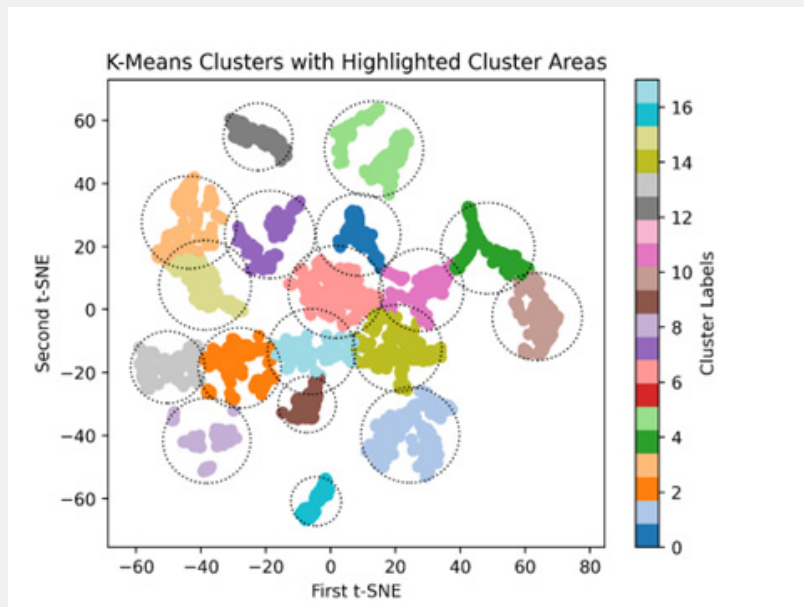
each species (HS= Homo sapiens, RN= Rattus norvegicus, MM= Mus musculus, LA= Loxodonta africana).

Cluster #	HS	LA	MM	RN	Total
1	49	40	114	20	223
2	36	5	48	25	114
3	71	83	22	4	180
4	75	50	48	10	183
5	72	26	92	19	209
6	71	60	72	49	252
7	65	99	131	21	316
8	15	14	79	31	139
9	69	21	100	23	213
10	2	18	45	26	91

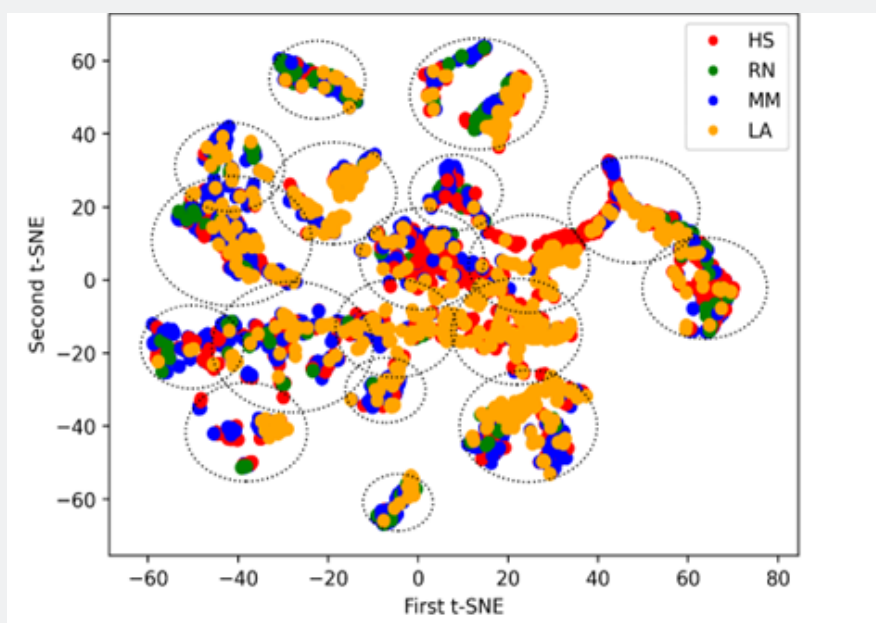
11	21	23	111	17	172
12	74	28	66	24	192
13	17	86	61	3	167
14	41	15	108	18	182
15	70	60	84	10	224
16	75	54	10	7	146
17	67	95	15	2	179
18	57	16	51	12	136



**Figure 1:** The figure shows the number of ORs of each species, indicated by abbreviations (HS= Homo sapiens, RN= Rattus norvegicus, MM= Mus musculus, LA= Loxodanta africana) The light blue line at the top of the graphs shows the total number of ORs per cluster. Individual numbers and comparative numbers per cluster are shown by the yellow, dark blue, orange and gray lines.



**Figure 2a:** The eighteen clusters of the 3318 olfactory receptor genes for genes from the four organisms. The color legends depict the cluster numbers. The clusters are demarcated by dotted circles.



**Figure 2b:** The eighteen clusters of the 3318 olfactory receptor genes showing the distribution of the genes for the four organisms in each cluster. The color legends show the genes for each species. The clusters are demarcated by dotted circles.

The number of ORs in each cluster ranged from 91 to 316. Cluster #7 had the largest number of ORs with 316 (65 [HS], 99 [LA], 131 [MM], and 21 [RN]), Cluster #10 had the fewest ORs with 91 (2 [HS], 18 [LA], 45 [MM], and 26 [RN]). The greatest number of ORs in each cluster were: #4 & 16 for humans, #7 for elephant, #7 for mouse, and #6 for rat. The fewest number of ORs in each cluster were: #10 for humans, #2 for elephant, #16 for mouse, and #17 for rat.

As might be concluded from (Figure 1), the mouse olfactory receptors number higher than the other three species and are therefore the highest represented in most clusters. There are a few discrepancies, however. The elephant ORs number the least in cluster #2, while higher than in mouse in clusters #3, #13, and #17. Mouse ORs are fewer in number in clusters #3, #16, and #17. The human ORs trend as is expected from the relative number of ORs when compared to mouse, rat and elephant, but are fewest in cluster #10 and considerably fewer than mouse and elephant ORs in cluster #13.

(Figures 2a&2b) show the distribution of the 18 clusters with dotted circles representing each cluster. (Figure 2a) illustrates the distribution of clusters with each cluster color-coded. (Figure 2b) shows the clusters with representations of the olfactory receptor genes for each organism (color-coded), each sequence is represented with a filled circle, following dimensionality reduction described in the Methods section. The position of each circle in this representation of the olfactory space was determined through a high dimensionality distance matrix following multiple sequence alignment.

## Discussion

In a previous paper [25], we showed that ORs and mouse and humans were found in the same clusters, though because of the larger number of functional ORs in mice, each cluster had a larger number of mouse ORs than human ORs. This study showed that each cluster is populated by olfactory receptor gene representations for human and mouse and there were no unique clusters composed of only human or only mouse ORs. In this work, while (Figure 1) and (Table 1) shows that there is variation in the OR populations of each cluster, the results of the previous study are largely confirmed. And this study considers two additional species whose olfactory repertoires are well characterized. While additional mammals would have to be included to definitively state this, it is more likely that the mammalian olfactory space is the similar.

A potential extension to this study is to repeat the assessments presented in this work using additional strategies—after increasing the cohort to include functional amino acid sequences for all mammalian ORs, assuming that a genome for that mammal is published.

Olfactory receptors are membrane bound sequences that can be seen as structurally three separate components—the N-termini and the extracellular loops, the transmembrane helical domain, and the intracellular loops and the C-termini (though there are exceptions to this sequence-structure paradigm). The upper part of the transmembrane helical domain contains a binding region that does not allow access to the interacting odorant with the cytoplasmic region of the cell. Namely, the helices 3 and 4 cross

to form a “cone” structure in the interior of the transmembrane bundle. We have previously shown through computational simulation studies that a putative transit path exists through which the odorant likely enters and exits (following OR activation), the binding region of the OR [26].

Hypothetical sequences consisting of only the binding region (the top half) and cytoplasmic side could be clustered independently to assess whether the clusters of the odorant-binding regions are different from the G-protein binding regions or different from those of the entire amino acid sequence of the olfactory receptor.

OdorDB (<https://ordb.biotech.ttu.edu/odordb>) is a companion database to ORDB and is a repository of odorant molecules that have been shown through experimental functional analysis to interact with (activate or inhibit) ORs. The process of deorphanizing ORs is challenging. These membranes bound proteins are difficult to express and purify [27-28]. OR-odorant interactions are considered promiscuous. ORs are known to interact with a wide range of odorants, whereas others are narrowly tuned. Odorants, on the other, hand are known to interact with different ORs [29]. Identifying rational panels of odorants which will establish a pattern of interaction between ORs, and odorant interaction behavior has yet to be established. Our methodology when extended will allow clusters comprising of sequences that represent the binding regions of ORs to be created. The next step would be to place odorants known to bind and activate ORs in the clusters containing these ORs. This would be a significant step towards identifying ORs that are likely to bind odorants with specific functional groups or electronic structural features [30].

## Conclusion

We have performed a comprehensive analysis of the distribution of functional olfactory receptor genes of four species: human, mouse, rat and elephant. This is with a view to determine and provide visual evidence of the distribution of these genes based on protein-sequence similarity in what we can identify as the olfactory space. Our studies reveal that the best resolved distribution of these gene sequences is in 18 clusters. Each cluster is populated. This leads to the primary conclusion that the olfactory space while varied is consistent for mammals. Additional work needs to be done to include more mammalian olfactory repertoires to confirm this.

## References

- Buck L, Axel R (1991) A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* 65(1): 175-187.
- Van Der Weyden MB, Armstrong RM, Gregory AT (2005) The 2005 Nobel Prize in physiology or medicine. *Med J Aust* 183(11/12): 612.
- US DOE Joint Genome Institute: Hawkins Trevor Initial sequencing and analysis of the human genome. *Nature* 409(6822): 860-921.
- Zozulya S, Echeverri F, Nguyen T (2001) The human olfactory receptor repertoire. *Genome Biol* 2(6): 1-12.
- Glusman G, Yanai I, Rubin I, Lancet D (2001) The complete human olfactory subgenome. *Genome Res* 11(5): 685-702.
- Niimura Y, Nei M (2003) Evolution of olfactory receptor genes in the human genome. *Proc Nat Acad Sci* 100(21): 12235-12240.
- Zhang X, Firestein S (2002) The olfactory receptor gene superfamily of the mouse. *Nature Neurosci* 5(2): 124-133.
- Young JM, Friedman C, Williams EM, Ross JA, Tonnes-Priddy L (2002) Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Human Mol Genet* 11(5): 535-546.
- Quignon P, Giraud M, Rimbault M, Lavigne P, Tacher S, et al. (2005) The dog and rat olfactory receptor repertoires. *Genome Biol* 6(10): 1-9.
- Niimura Y, Matsui A, Touhara K (2014) Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. *Genome Res* 24(9): 1485-1496.
- Takahashi LK (2014) Olfactory systems and neural circuits that modulate predator odor fear. *Front Behav Neurosci* 8: 72.
- Bombail V (2019) Perception and emotions: on the relationship between stress and olfaction. *Appl Animal Behav Sci* 212: 98-108.
- German JB, Yeritzian C, Tolstoguzov VB, (2007) Olfaction, where nutrition, memory and immunity intersect. In *Flavours and Fragrances: Chemistry, Bioprocessing and Sustainability*, Berlin, Heidelberg: Springer Berlin Heidelberg 25-41.
- Majid A, Kruspe N (2018) Hunter-gatherer olfaction is special. *Curr Biol* 28(3): 409-413.
- Sharon D, Glusman G, Pilpel Y, Khen M, Gruetzner F, et al. (1999) Primate evolution of an olfactory receptor cluster: diversification by gene conversion and recent emergence of pseudogenes. *Genomics* 61(1): 24-36.
- Hasin Y, Olender T, Khen M, Gonzaga-Jauregui C, Kim PM, et al. (2008) High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution. *Plos Genetics* 4(11): e1000249.
- Inan FA (2023) Geographic Differences in Genomic Variations of Human Olfactory Genes: A Feasibility Study. Capstone Report. M.S. in Biotechnology. Center for Biotechnology and Genomics. Texas Tech University, Lubbock, Texas, USA.
- Nadkarni PM, Marengo L, Chen R, Skoufos E, Shepherd G (1999) Organization of heterogeneous scientific data using the EAV/CR representation. *J Am Med Inform Assoc* 6(6): 478-493.
- Marengo L, Tosches N, Crasto C, Shepherd G, Miller PL, et al. (2003) Achieving evolvable Web-database bioscience applications using the EAV/CR framework: recent advances. *J Am Med Inform Assoc* 10(5): 444-453.
- Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, et al. (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res* 40(W1): W597-W603.
- Sievers F, Higgins DG (2014) Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* 1079: 105-116.
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7(1): 539.
- Li W, Cerise JE, Yang Y, Han H (2017) Application of t-SNE to human genetic data. *J Bioinform Comput Biol* 15(4): 1750017.
- Steinley D (2006) K-means clustering: a half-century synthesis. *Br J Math Stat Psychol* 59(1): 1-34.
- Man O, Willhite DC, Crasto CJ, Shepherd GM, Gilad Y (2007) A framework for exploring functional variability in olfactory receptor genes. *Plos One* 2(8): e682.

26. Lai PC, Singer MS, Crasto CJ (2005) Structural activation pathways from dynamic olfactory receptor-odorant interactions. *Chem Senses* 30(9): 781-792.
27. Kiefer, Krieger H, Olszewski J, Von Heijne JD, Prestwich G, et al. (1996) Expression of an olfactory receptor in *Escherichia coli*: purification, reconstitution, and ligand binding. *Biochem* 35(50): 16077-16084.
28. Touhara K (2007) Deorphanizing vertebrate olfactory receptors: recent advances in odorant-response assays. *Neurochem Int* 51(2-4): 132-139.
29. Krautwurst D (2008) Human olfactory receptor families and their odorants. *Chem Biodiversity* 5(6): 842-852.
30. Bhayani H, Thilakaratne R, Perera N, Crasto C (2023) Exploring the Molecularity of the Odor and Taste Perceptions of "Brown": A Computational Approach. In *Proc Int Confer New Trend Appl Sci* 1: 98-102.



This work is licensed under Creative Commons Attribution 4.0 License

DOI:[10.19080/JOJWB.2024.05.555654](https://doi.org/10.19080/JOJWB.2024.05.555654)

**Your next submission with Juniper Publishers  
will reach you the below assets**

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats

**( Pdf, E-pub, Full Text, Audio )**

- Unceasing customer service

**Track the below URL for one-step submission**

<https://juniperpublishers.com/online-submission.php>