

## HORDE: Comprehensive Resource for Olfactory Receptor Genomics

Tsviya Olender, Noam Nativ, and Doron Lancet

### Abstract

Olfactory receptors (ORs) constitute the largest gene family in the mammalian genome. The existence of these proteins underlies the nature of, and variability in, odorant perception. The Human Olfactory Receptor Data Explorer (HORDE, <http://genome.weizmann.ac.il/horde/>) is a free online resource, which presents a complete compendium of all OR genes and pseudogenes in the genome of human and four other vertebrates. HORDE includes three parts: (1) an automated pipeline, which mines OR gene and pseudogene sequences out of complete genomes, and generates gene symbols based on sequence similarity; (2) a card generator that obtains and displays annotative information on individual ORs retrieved from external databases and relevant studies; and (3) a search engine that allows user retrieval of OR information. For human ORs, HORDE specifically addresses the universe of interindividual variation, as obtained from several sources, including whole genome sequences made possible by next-generation sequencing. This encompasses single nucleotide polymorphisms (SNP) and copy number variation (CNV), including deleterious mutational events. HORDE also hosts a number of tools designed specifically to assist in the study of OR evolution and function. In this chapter, we describe the status of HORDE (build #43). We also discuss plans for future enhancements and a road map for HORDE to become a better community-based bioinformatics tool. We highlight HORDE's role as a major research tool in the study of an expanding cohort of OR repertoires.

**Key words** Olfactory receptors, Database, SNP, Copy number variation, Computational data-mining

---

### 1 Introduction

Olfaction is initiated following the binding of an odorant ligand to an olfactory receptor (OR) protein, which is expressed in the ciliary membrane of olfactory neurons (1–3). ORs, the largest gene superfamily in the mammalian genome, are composed of seven transmembrane domain proteins, members of the G-(Protein)-coupled protein receptors (GPCRs) hyperfamily. A typical mammalian genome contains 1,000–1,500 OR genes, evidence for the essential role of the olfactory system to survival of many species (4–6). HORDE, the Human Olfactory Receptor Data

Exploratorium, is a resource consisting of algorithmic and data storage processes that mine, annotate, and integrate complete repertoires of vertebrate OR genes. It is mainly designed to provide insight into the evolution, structure, and function of the complete human OR repertoire. Its main mission is to facilitate the navigation and scrutiny of such large gene repertoires. Another OR database is ORDB (7) (see Chapter 1), which is broader in scope, encompassing all chemosensory receptors (olfactory, vomeronasal, and gustatory receptors), but contains less information on individual human ORs. GPCRDB (8), which extracts and stores GPCR information only from public genome sources, has an OR gene section.

HORDE was established in 2000, following the publication of the first draft of the human genome (9–11). It presented for the first time a complete human OR repertoire, the result of computational data mining. Over time, HORDE has been significantly expanded and improved (12, 13), with a mission to present a complete and nonredundant digital compendium of human ORs. OR repertoires of several other vertebrate genomes have also been mined and annotated by HORDE and added to the database.

The OR repertoires of many species are replete with pseudogenes, the result of frame-disrupting mutations that spread and got fixed in the population (14). Such a process is rationalized by the functional redundancy of OR proteins, whereby each odorant is recognized by several ORs (15–17). The presence of pseudogenes is particularly noticeable in primates, which apparently depend less on olfaction for their survival, culminating in the inactivation of ~55 % of all ORs in humans. These pseudogenes are not usually identified by the standard whole genome annotation pipelines. Notably, OR pseudogenes are not annotated even by databases devoted to pseudogenes (18), because they are not processed pseudogenes (like those formed by RNA retroposition). The HORDE specialized computational pipeline allows it to be a unique resource that includes all gene and pseudogenes for a given genome.

Although ORs have been investigated for nearly two decades (2, 19), many relevant questions have not been fully answered. One example is the genetic basis of the widespread phenotypic diversity in human olfactory faculties. This is believed to be attributable, at least in part, to genomic variations in the OR genes (6, 20, 21). Relevant studies, therefore, require an accurate and comprehensive account of the genes and their genetic variations. HORDE also contains information on expression (obtained from ESTs or microarray data), as well as genomic location and clustering of ORs, and this information is relevant to several open questions. Among these are the control of gene expression in the olfactory system, including epithelial zone-specific expression as well as expression with locus and allele

exclusion (22, 23), ectopic expression in non olfactory tissues (24–26), such as sperm (27), and biased expression of certain OR genes and pseudogenes (e.g., (28)).

In this chapter we describe in detail the information presented in HORDE and its use in studies of the OR superfamily. We give the full details on the annotation pipeline, HORDE's classification methods, and the analysis routines. We also describe database design, and the tools for data retrieval.

---

## 2 Data-Mining and Database Design

HORDE is committed to retrieving, processing, and archiving of the entire nonredundant list of all human ORs. We have developed an automatic pipeline (Fig. 1) that can be reapplied whenever a new version of the human genome assembly is released. This procedure, which is also used to decipher the OR repertoires of other species, is facilitated because the coding region of the OR gene is encoded by a single exon, and because the OR protein sequences are mutually conserved and distinguishable from non-OR GPCR proteins (10). Similar approaches, with minor modifications, have been later utilized by others (29–32).

This pipeline was applied to decipher the OR repertoire of the human, dog (33), opossum (34), platypus (15), lizard (35), and zebra finch (35), and with minor changes, to define the VIR repertoire of the platypus (15).

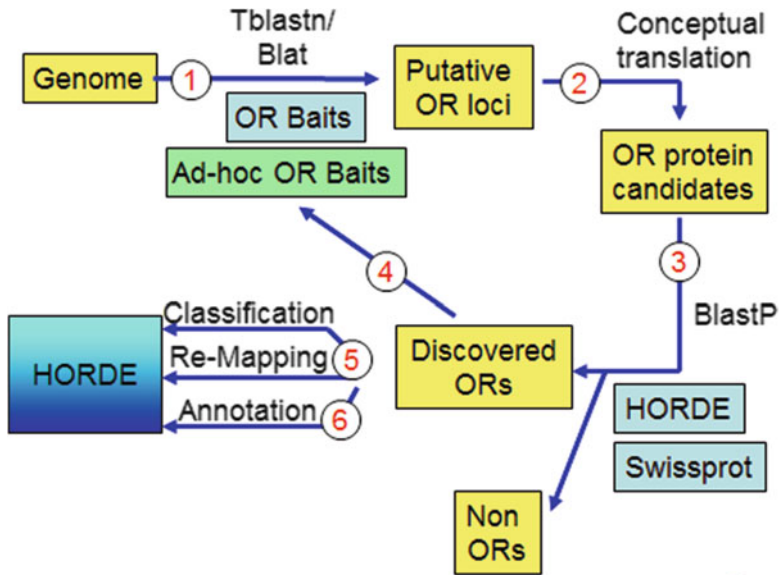
The result of our data-mining efforts is a database where information has been currently stored in a flat file format. A user search and retrieval interface with an interface written in PERL (Practical Extraction and Report Language). A new relational MySQL database is now under development (See Note 2). The new database design (Fig. 2) will be equipped with the open source CakePHP framework (<http://cakephp.org/>) that uses a built-in Object-Relational Mapping (ORM), thus providing a simple API (application program interface) for data storage and retrieval.

---

## 3 Description of the HORDE Database

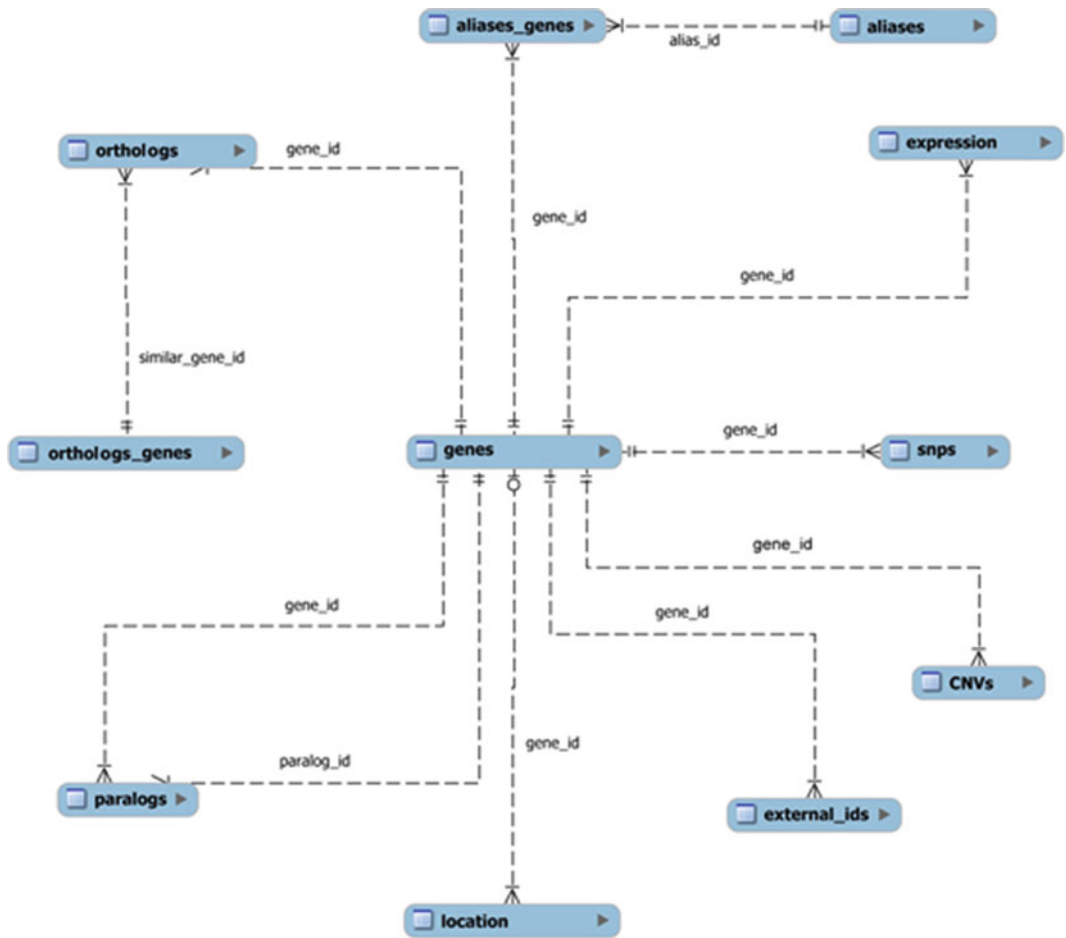
### 3.1 *The OR Repertoires of Human and Other Species*

Table 1 shows the OR counts within HORDE (build #43) for genes and pseudogenes in the species included. For all the species, the OR superfamily is classified by sequence similarity using BLASTP into 18 families and 454 subfamilies (not all species have all OR families and subfamilies) (13). The details of the classification method were described in (10, 12, 33, 36). Briefly, a new gene is classified into the same subfamily as that of its best hit if it shows  $\geq 60\%$  protein sequence identity to such target. It is classified into a new subfamily of the same family as that of its best hit if it shows



**Fig. 1** The steps of the OR mining pipeline (see Note 1). (1) genomic screening by BLAT (62) or TBLASTN (63) with a set of “OR baits” to discover genomic putative OR loci, locations suspected to harbor OR genes. TBLASTN is preferable when mining a new genome, the faster BLAT procedure is sufficient for updating an existing repertoire. As “OR baits” we use a set of about 500 OR representatives coming from different OR subfamilies, and different vertebrates species, designed to maximally cover the OR sequence space (<http://genome.weizmann.ac.il/horde/DataMining.htm>). A locus is suspected as “putative OR” if the in TBLASTN  $E < 10^{-3}$  over at least 50 amino acids, or if it constitutes a BLAT search hit. (2) BioPerl is used for conceptual translation and the ORFs of 290–350 amino acids are accepted as - ORs. ORFs outside this range are suspected as OR pseudogenes and are therefore subjected to FASTY analysis to generate a “corrected” OR (64). (3) The resulting “OR protein candidates” are classified as ORs or non-ORs by BLASTP comparison against the current version of HORDE and against all Swissprot proteins. OR candidates which their sequence is at least 40 % identical over at least 100 amino acids to another OR are considered as OR genes (10). Borderline cases with <40 % protein identity to the best OR, with another GPCR being the best hit in EXPASY (formerly, SwissProt), are classified as non-ORs. (4) Steps 1–3 are repeated using newly discovered ORs as baits (“Ad-hoc OR baits”), until no novel hits are detected. (5) OR genes are remapped to the genome; they are classified into families and subfamilies; a unique symbol is assigned to each gene (see text). (6) Additional annotation is integrated into the database from external sources. The external sources currently accessed are GeneBank, dbSNP, ORDB (7), DGV (43), and BioGPS (47). Updates necessary after the publication of a new genome version for a species already included in HORDE are performed as follows: All ORs of this species undergo an update of genomic coordinates to those of the new version, using BLAT. Then the mining pipeline is reapplied; ORs already present are identified by their genomic coordinates, and newly discovered ones are added to HORDE and annotated

$\geq 40$  %, but  $\leq 60$  % protein sequence identity to such target. In cases where the similarity to the best hit is lower than  $\leq 40$  %, the gene is candidate to define a new family, but such a definition is carried out with caution, by a more elaborate procedure that involves phylogenetic and principal component analyses.

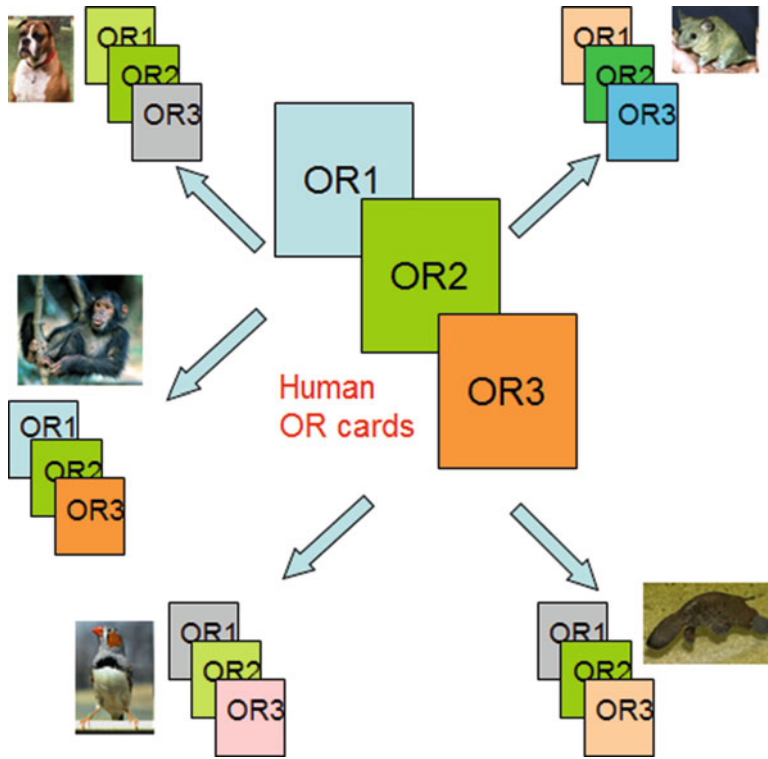


**Fig. 2** HORDE version #43 design in a relational model. The *boxes* represent the database tables. The *dashed arrows*, labeled with the foreign key fields, show the relationships between the tables

**Table 1**  
**Summary of the OR repertoires listed in HORDE**

Species	Reference	Number of intact ORs	Number of pseudogenes
Human	(13)	390 (387)	468 (415)
Chimpanzee	(16)	397 (380)	693 (414)
Dog	(33)	859	315
Opossum	(33)	613	898
Platypus	(15)	350	404

Also given in parenthesis, the gene count of the human and chimpanzee OR repertoires from the study of (65), for comparison (see Note 4)



**Fig. 3** HORDE is a human-centric database but includes OR repertoires from other vertebrates. The information is presented to the user via an “OR card” for all species. The information for human ORs is extensive: it is integrated with external relevant sources. The human OR card contains links to the ortholog OR cards in other species

Based on the classification, HORDE provides a HUGO-approved systematic nomenclature (<http://www.gene.ucl.ac.uk/nomenclature/genefamily.shtml>) that affords an instant guide to the position of a gene in a phylogenetic tree (36). The nomenclature system consists of a root “OR,” followed by the family numeral, the subfamily letter(s), and a numeral representing the individual gene within the subfamily. For example, OR3A1 is an OR gene of family 3, subfamily A, and OR7E12P is an OR pseudo-gene of family 7 subfamily E. This nomenclature system for human ORs is now widely.

### 3.2 The Human OR Card

A key feature in HORDE is a “card” for every human OR gene (Figs. 3 and 4a), a concept adopted from the GeneCards compendium of human genes (9, 37). The OR card includes rich annotation relevant to this gene, including links to internal tools as well as to appropriate external databases. A description of the different sections of the human OR card is provided below.

#### 3.2.1 Gene Nomenclature and Phylogeny

This section contains information regarding the classification of the OR, its aliases and its functionality (being intact or pseudo-gene). OR genes with in-frame stop codons or frame shifts are

**HORDE** Horde #42-404: human OR11H7P

### Additional Exons\*

\* information was derived from ESTs and mRNAs, may be incomplete

### Sequence information

Nucleotide sequence (G+C = 45.4%, CpG CV = 0.10)

>0911379 (R309E#42: 404)  
ATGAATACCTCCAGATGATCATCTGTGACGCAATTTGTGTGTTGGGG  
CTCTGGCTGCTGGGAAATTCAGATGCTCTTTTTCTCAATGATTTGTGTGG  
TCTACATCTTCTACTCTGACGTGGGAATCTGGCCATCATCTGTGCAATTGAGG  
TGGGAGCATCTGACCTGACCCGCTATGTACGTGCTCTCAAGCAGATCTGCT  
CTTCTCAGAGATCTTGATATGTGACCTGCGACATGCTCCCAAGCATCTGTGTAA  
ATTTTTCTCTCAAAAACCTAAGACCATATGATCTCTTGATGATTTTGATCAGC  
TGACCATCTCTCTCTCTCTCTGGGACCATCACTGATGCTCTCTCTCTCTCTCT  
CATGGCTATGATCTGGTAAGCTGGGACATCTGACCCCATGCACTATGCTCT

**Fig. 4 (a)** An example of human OR card. Note, that the OR presented, OR11H7P, is a Segregating PseudoGene, due to a SNP, which alters a frame disrupting stop codon into glutamine. A recent genomic study reported on an association between this SNP and isovaleric acid sensitivity (60). **(b)** An example of nonhuman OR card. The canine's ortholog of OR11H7P



b

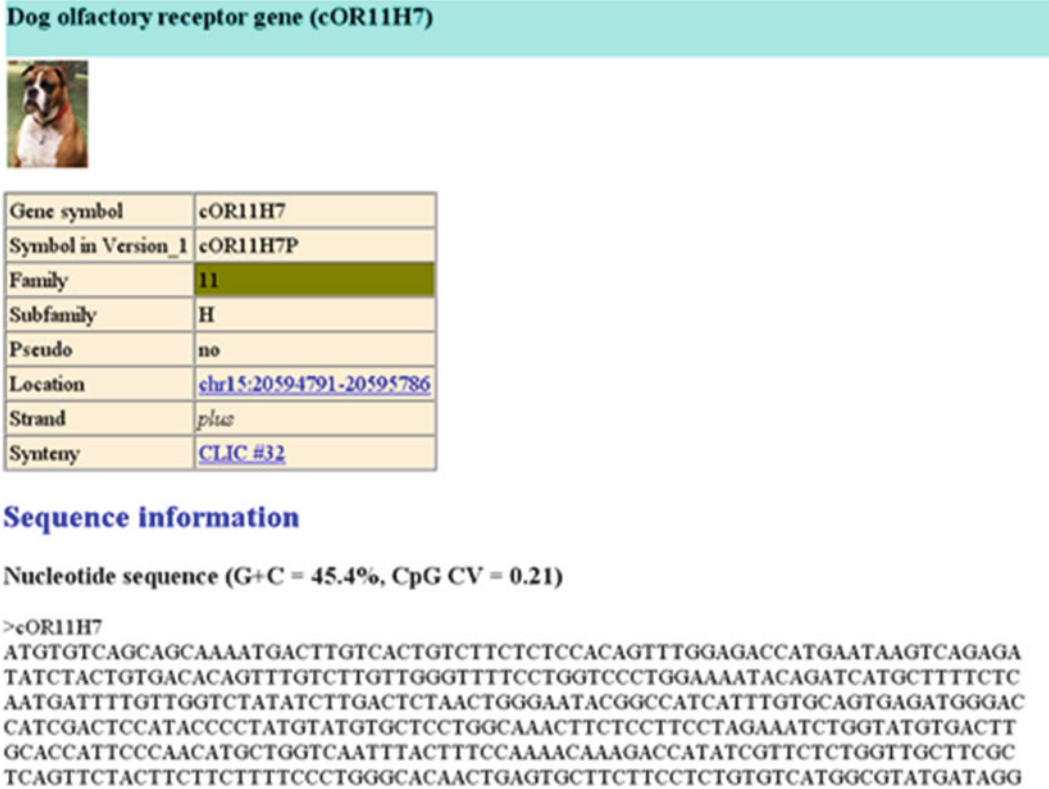


Fig. 4 (continued)

classified as pseudogenes; the rest are classified as intact genes. Genes with both intact and disrupted alleles due to specific SNP that leads to allele inactivation are stored in a special category are (38), and are classified as Segregating PseudoGenes (SPGs). We determine a “pseudogene probability” score, which assesses the probability of an intact OR gene to encode a functional protein (Subheading 4.3.1).

3.2.2 Mapping

The mapping section provides a global overview of the genomic organization of the OR superfamily in the mammals genome. Thus, information is given on the genomic location of the gene, the OR cluster (Fig. 5) and on the conservation of the cluster organization in five mammalian genomes (CLusters In Conservation, CLICs) (34).

The information about the cluster is linked to a “cluster viewer,” a special feature of HORDE, which facilitates browsing of the ORs organization along human chromosomes (Subheading 4.3.3 provides a bit more details).

CLICs information is linked to a dedicated Web page within HORDE (<http://genome.weizmann.ac.il/horde/CLICs/>), which presents the syntenic clusters in other species (Subheading 4.3.2).



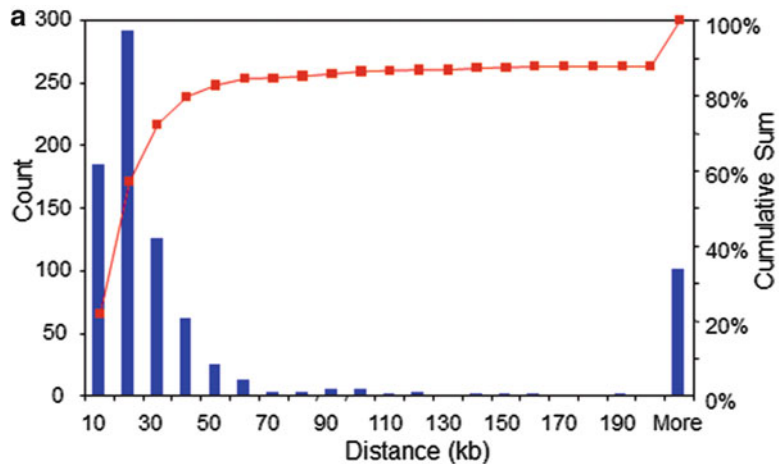
### 3.2.3 OR Paralogs and Orthologs

The OR superfamily have been subjected to frequent gene recombination, gene conversion, duplication, and translocation events (14, 39). These events make the identification of orthologs and paralogs a nontrivial task. By definition, HORDE does not attempt to identify orthologs and paralogs authoritatively, but rather supplies for the documented gene (probe) the top BLASTP hits in the human repertoire and in some other mammals (Fig. 6). If the best hit in a given species has  $X$  % identity to the probe, HORDE presents all results that have identity scores  $>(X-5)$  % to the probe.

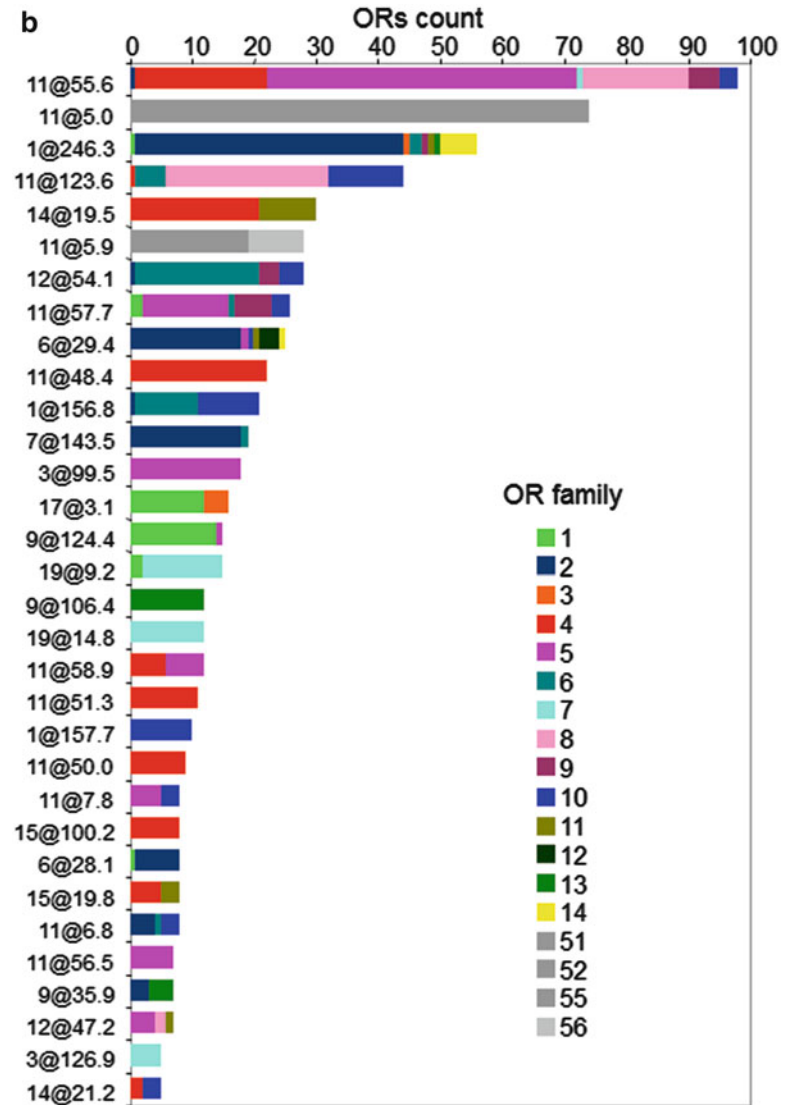
### 3.2.4 Genomic Polymorphisms

Human ORs are enriched in genomic variations (40, 41, 66), a result of gene redundancy that also leads to the gradual evolutionary diminution of the human OR repertoire. HORDE aims to contain a comprehensive collection of genomic variations in the OR genes, presented in the genomic polymorphisms section.

SNPs and small indels are extracted from dbSNP (42), the 1,000 genomes project (<http://www.1000genomes.org/page.php>), and an in-house next generation sequencing of 100 genes. CNVs are extracted from the Database of Genomic Variants (DGV, <http://projects.tcag.ca/variation/>) (43), MoDIL (<http://compbio.cs.toronto.edu/modil/>) (44) and dbSNP (build 131), as well as from a published paper (45).



**Fig. 5** Genomic organization of the human OR repertoire. (a) ORs are disposed in the genome in compact clusters, with a mean distance of  $20 \pm 15$  kb. The bars show the distribution of the distances of each two consecutive ORs. The cumulative sum is represented by a line dotted with squares. (b) HORDE divides the human OR repertoire into 136 clusters ranging from one (single ton) to 97 ORs per cluster. The largest 20 clusters are shown, including their OR families composition

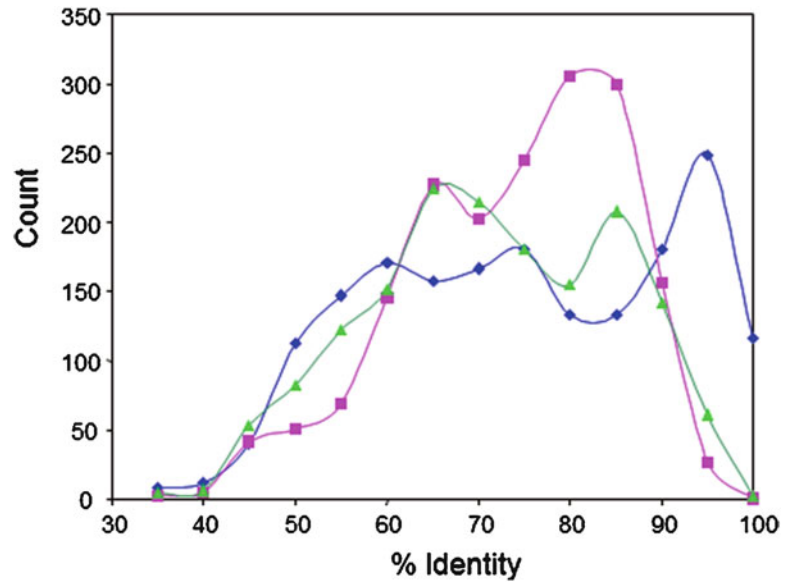


**Fig. 5** (continued)

All genomic variations in HORDE (Table 2) are collected based on genomic overlap with the OR coordinates. SNP data is extracted using the UCSC “Table browser” tool (<http://genome.ucsc.edu/cgi-bin/hgTables?command=start>) and further analyzed by our Perl script that interprets their effect on the protein (i.e., whether they are synonymous or non-synonymous). SNPs with a *map weight* > 1 (a parameter indicating nonunique genome mapping (46)), are shown with multiple map targets. (see Note 3)

3.2.5 Expression  
Evidence  
and Additional Exons

ORs are expressed in the olfactory epithelium, but also ectopically in many other tissues (24–26). The expression evidence section provides information of gene expression extracted from BioGPS



**Fig. 6** A distribution of the % identity values of the human most similar ORs in mouse (*line with squares*), dog (*line with triangles*) and human (*line with diamonds*). Such a distribution can be used to study orthology and paralogy relationship

**Table 2**

**A summary of all genomic polymorphic events in HORDE (66)**

ORs	SNPs	In/Dels (<1,000 bp)	CNVs	CNVs with deletion allele
Intact	5,362	204	291	98
Pseudo	5,591	452	360	133

In/Dels=insertions or deletions

DNA arrays (47, 48) and *in silico* data mining of ESTs and mRNA (12, 25). The full details of the data mining and analysis of ESTs and mRNAs are described in (25) and (12).

The mRNA that encodes ORs typically includes a single downstream exon that contains the entire ~900 bp coding region, and encompasses ~1–2 kb 3'-UTR. The upstream part of the transcript often has several short 5' untranslated exons, with evidence for alternative splicing, both by standard intron removal and by in-exon splicing (49–52). Unfortunately, the untranslated regions of the OR genes are not conserved and cannot be readily identified by current gene prediction programs or by HORDE's pipeline. By analyzing spliced ESTs, we defined and added to the database information on 388 untranslated exons belonging to 149 ORs (12).

### 3.2.6 Protein Domains and Features

This section contains functional annotation of the OR protein which includes the seven transmembrane domains, positions suspected as the putative odorant binding site residues, termed the complementary determining residues (CDRs) (53), a conserved N-glycosylation site, and two conserved cysteines that are predicted to form a disulfide bond between extracellular loops 1 and 2 (53, 54). The annotation is performed by aligning each OR protein using the profile mode of CLUSTALW to the highly curated and annotated alignment published in (53).

Currently (build #43), the “protein domains and features” section is available to the users through a link from the main OR card to a dedicated html page.

### 3.3 Nonhuman OR Cards

Information about other vertebrate ORs is also presented by a “card.” These cards, however, contain less information than that for humans. The card typically includes the following: the OR classification, genomic location, and sequences (Fig. 4b). The dog and opossum OR cards contain links to their relevant CLIC.

---

## 4 HORDE Query Tools

HORDE is equipped with several search tools, allowing a text-free search as well as sequence search.

### 4.1 Sequence Search Tool

The BLAST sequence search tool provides access to a large collection of OR repertoires that we have included within HORDE. Besides the OR repertoires listed in Table 1, library access is provided to the OR repertoires of the mouse (31, 55), rat (56) and chicken (extracted from GenBank). Another library, named the “zoo library,” contains 2,235 OR partial sequences we have mined from 47 vertebrates collected from GenBank.

### 4.2 Text Search Tools

HORDE offers the user two text-search tools. The first is a “free text” tool. For example, a search with the text “Segregating PseudoGene” will result in a list of all SPGs in the database. A search using the term “rs379147,” will return a link to OR1E1, the gene that contains this SNP. The second tool allows the user to select for a group of ORs based on a shared chromosomal location, or classification. For example, under the option *Select all ORs from:* inserting the number “1” for the *family*, results in a list of all ORs from family 1. The nucleotide or protein sequence can be then retrieved for all or selected ORs in this list.

### 4.3 HORDE Analysis Tools

#### 4.3.1 CORP (<http://genome.weizmann.ac.il/horde/CORP/>)

The probabilistic Classifier for Olfactory Receptor Pseudogenes (CORP) algorithm (57) computes the probability of an OR gene with intact open reading frame to encode a nonfunctional protein (i.e., a pseudogene) by examining the deviation of its protein sequences from the OR functionally crucial consensus. It accepts as an input an OR protein sequence and returns the pseudogene probability score of the gene.

#### 4.3.2 CLICs (<http://genome.weizmann.ac.il/horde/CLICs/>)

CLIC (Cluster In Conservation) is a tool for genome-wide definition of genomic gene clusters conserved in multiple species (34). The CLIC algorithm was applied to the ORs of five mammals. 48 multispecies CLICs were discovered, covering more than 90 % of all OR genes. The CLIC Web page contains a list of all CLICs and full details of the ORs participating in each CLIC from all compared species.

#### 4.3.3 The Cluster Viewer

This presents the cluster details (location, size, and OR family composition) and the OR cluster members. A link to the closest OR clusters allows the user to browse the OR clusters along a given chromosome. A cluster is a group of ORs with maximal intergenic distance of a 100 kb (Fig. 5).

---

## 5 Notes

1. Although HORDE data-mining routine was successfully applied to several vertebrate genomes, it might be argued that the use of BLAST search is too conservative and the result is biased toward genes showing similarity to the “OR baits,” potentially missing some *bona fide* OR genes. For this reason, we make sure that we perform the data mining procedure until known GPCR proteins are identified.
2. HORDE is currently stored in a flat-file format equipped with Perl-CGI interface. For the next version (August 2012), all the data stored into HORDE will be ported into a relational database. Leveraging the flexibility and complexity that such a data-infrastructure can contribute, it will allow complex user-queries and integration with other sources.
3. The information on the complete OR gene structure (Subheading 3.2.5) is partial and preliminary. A new mining procedure is now under development and will be based on the AceView analysis (58), which performed the same analysis but on a whole genome scale and thus is more accurate.
4. The increasing number of sequenced individual genomes will dramatically increase our knowledge of the known genomic variations in the human ORs. Of special relevance are deleterious

variations, nonsense SNPs, which cause premature stop codon, null allele CNVs (deletions) leading to OR inactivation (45, 59), or substitution in a highly conserved amino acid. Such polymorphisms might lead to phenotypic differences in olfactory acuity and perception, as was demonstrated in the study of (60) and (61). One of HORDE's challenges is to present a continuously updated compendium of genomic variations, thus providing the infrastructure for future genomic genetic association studies of OR genotypes to olfactory phenotypes.

5. A companion database is now available for olfactory auxiliary genes related to transduction and sensory neuronal integrity (<http://genome.weizmann.ac.il/GOSdb/>, REF: PMID: 22936402).

---

## Acknowledgements

This research was supported by the Israel Science Foundation Heritage fund, LifeMap Sciences Inc. California (USA), as well as the Crown Human Genome Center and the Nella and Leon Benoziyo Center for Neurological Diseases at the Weizmann Institute of Science.

## References

1. Touhara K, Vosshall LB (2009) Sensing odorants and pheromones with chemosensory receptors. *Annu Rev Physiol* 71:307–332
2. Kaupp UB (2010) Olfactory signalling in vertebrates and insects: differences and commonalities. *Nat Rev Neurosci* 11:188–200
3. Rodriguez-Gil DJ, Treloar HB, Miller A et al (2009) Onset of odorant receptors. *Ann N Y Acad Sci* 1170:18–20
4. Nei M, Niimura Y, Nozawa M (2008) The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet* 9:951–963
5. Rouquier S, Giorgi D (2007) Olfactory receptor gene repertoires in mammals. *Mutat Res* 616:95–102
6. Kambere MB, Lane RP (2007) Co-regulation of a large and rapidly evolving repertoire of odorant receptor genes. *BMC Neurosci* 8(Suppl 3):S2
7. Crasto C, Marenco L, Miller P et al (2002) Olfactory receptor database: a metadata-driven automated population from sources of gene and protein sequences. *Nucleic Acids Res* 30:354–360
8. Horn F, Bettler E, Oliveira L et al (2003) GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res* 31:294–297
9. Safran M, Chalifa-Caspi V, Shmueli O et al (2003) Human gene-centric databases at the Weizmann Institute of Science: GeneCards, UDB, CroW 21 and HORDE. *Nucleic Acids Res* 31:142–146
10. Glusman G, Yanai I, Rubin I et al (2001) The complete human olfactory subgenome. *Genome Res* 11:685–702
11. Fuchs T, Glusman G, Horn-Saban S et al (2001) The human olfactory subgenome: from sequence to structure and evolution. *Hum Genet* 108:1–13
12. Olender T, Feldmesser E, Atarot T et al (2004) The olfactory receptor universe—from whole genome analysis to structure and evolution. *Genet Mol Res* 3:545–553
13. Olender T, Lancet D, Nebert DW (2008) Update on the olfactory receptor (OR) gene superfamily. *Hum Genomics* 3:87–97
14. Niimura Y (2009) Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environments and genomic contents. *Hum Genomics* 4:107–118
15. Warren WC, Hillier LW, Marshall Graves JA et al (2008) Genome analysis of the platypus

- reveals unique signatures of evolution. *Nature* 453:175–183
16. Gilad Y, Man O, Glusman G (2005) A comparison of the human and chimpanzee olfactory receptor gene repertoires. *Genome Res* 15:224–230
  17. Kishida T, Kubota S, Shirayama Y et al (2007) The olfactory receptor gene repertoires in secondary-adapted marine vertebrates: evidence for reduction of the functional proportions in cetaceans. *Biol Lett* 3:428–430
  18. Karro JE, Yan Y, Zheng D et al (2007) Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. *Nucleic Acids Res* 35:D55–D60
  19. Mombaerts P (2004) Genes and ligands for odorant, vomeronasal and taste receptors. *Nat Rev Neurosci* 5:263–278
  20. Hasin-Brumshtein Y, Lancet D, Olender T (2009) Human olfaction: from genomic variation to phenotypic diversity. *Trends Genet* 25:178–184
  21. Menashe I, Lancet D (2006) Variations in the human olfactory receptor pathway. *Cell Mol Life Sci* 63:1485–1493
  22. Nguyen MQ, Zhou Z, Marks CA et al (2007) Prominent roles for odorant receptor coding sequences in allelic exclusion. *Cell* 131:1009–1017
  23. Imai T, Sakano H (2009) Odorant receptor gene choice and axonal projection in the mouse olfactory system. *Results Probl Cell Differ* 47:57–75
  24. De la Cruz O, Blekhman R, Zhang X et al (2009) A signature of evolutionary constraint on a subset of ectopically expressed olfactory receptor genes. *Mol Biol Evol* 26:491–494
  25. Feldmesser E, Olender T, Khen M et al (2006) Widespread ectopic expression of olfactory receptor genes. *BMC Genomics* 7:121
  26. Zhang X, De la Cruz O, Pinto JM et al (2007) Characterizing the expression of the human olfactory receptor gene family using a novel DNA microarray. *Genome Biol* 8:R86
  27. Fukuda N, Yomogida K, Okabe M et al (2004) Functional characterization of a mouse testicular olfactory receptor and its role in chemosensing and in regulation of sperm motility. *J Cell Sci* 117:5835–5845
  28. Xu LL, Stackhouse BG, Florence K et al (2000) PSGR, a novel prostate-specific gene with homology to a G protein-coupled receptor, is overexpressed in prostate cancer. *Cancer Res* 60:6568–6572
  29. Zhang X, Firestein S (2009) Genomics of olfactory receptors. *Results Probl Cell Differ* 47:25–36
  30. Young JM, Trask BJ (2002) The sense of smell: genomics of vertebrate odorant receptors. *Hum Mol Genet* 11:1153–1160
  31. Young JM, Friedman C, Williams EM et al (2002) Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Hum Mol Genet* 11:535–546
  32. Niimura Y, Nei M (2005) Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proc Natl Acad Sci USA* 102:6039–6044
  33. Olender T, Fuchs T, Linhart C et al (2004) The canine olfactory subgenome. *Genomics* 83:361–372
  34. Aloni R, Olender T, Lancet D (2006) Ancient genomic architecture for mammalian olfactory receptor clusters. *Genome Biol* 7:R88
  35. Warren WC, Clayton DF, Ellegren H et al (2010) The genome of a songbird. *Nature* 464:757–762
  36. Glusman G, Bahar A, Sharon D et al (2000) The olfactory receptor gene superfamily: data mining, classification and nomenclature. *Mamm Genome* 11:1016–1023
  37. Safran M, Dalah I, Alexander J et al (2010) GeneCards Version 3: the human gene integrator. *Database (Oxford)* 2010:baq020
  38. Menashe I, Man O, Lancet D et al (2003) Different noses for different people. *Nat Genet* 34:143–144
  39. Lapidot M, Pilpel Y, Gilad Y et al (2001) Mouse-human orthology relationships in an olfactory receptor gene cluster. *Genomics* 71:296–306
  40. Alonso S, Lopez S, Izagirre N et al (2008) Overdominance in the human genome and olfactory receptor activity. *Mol Biol Evol* 25:997–1001
  41. Nozawa M, Kawahara Y, Nei M (2007) Genomic drift and copy number variation of sensory receptor genes in humans. *Proc Natl Acad Sci* 104:20421–20426
  42. Day IN (2010) dbSNP in the detail and copy number complexities. *Hum Mutat* 31:2–4
  43. Zhang J, Feuk L, Duggan GE et al (2006) Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. *Cytogenet Genome Res* 115:205–214
  44. Lee S, Hormozdiari F, Alkan C et al (2009) MoDIL: detecting small indels from clone-end sequencing with mixtures of distributions. *Nat Methods* 6:473–474
  45. Hasin Y, Olender T, Khen M et al (2008) High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution. *PLoS Genet* 4:e1000249



46. [http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=helpsnpfq&part=Build.The\\_dbSNP\\_Mapping\\_Process#Build.Mapping\\_Weight](http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=helpsnpfq&part=Build.The_dbSNP_Mapping_Process#Build.Mapping_Weight)
47. Wu C, Orozco C, Boyer J et al (2009) BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol* 10:R130
48. Su AI, Wiltshire T, Batalov S et al (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* 101:6062–6067
49. Fleischer J, Breer H, Strotmann J (2009) Mammalian olfactory receptors. *Front Cell Neurosci* 3:9
50. Young JM, Shykind BM, Lane RP et al (2003) Odorant receptor expressed sequence tags demonstrate olfactory expression of over 400 genes, extensive alternate splicing and unequal expression levels. *Genome Biol* 4:R71
51. Michaloski JS, Galante PA, Malnic B (2006) Identification of potential regulatory motifs in odorant receptor genes by analysis of promoter sequences. *Genome Res* 16:1091–1098
52. Sosinsky A, Glusman G, Lancet D (2000) The genomic structure of human olfactory receptor genes. *Genomics* 70:49–61
53. Man O, Gilad Y, Lancet D (2004) Prediction of the odorant binding site of olfactory receptor proteins by human-mouse comparisons. *Protein Sci* 13:240–254
54. Menon ST, Han M, Sakmar TP (2001) Rhodopsin: structural basis of molecular physiology. *Physiol Rev* 81:1659–1688
55. Zhang X, Firestein S (2002) The olfactory receptor gene superfamily of the mouse. *Nat Neurosci* 5:124–133
56. Gibbs RA, Weinstock GM, Metzker ML et al (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428:493–521
57. Menashe I, Aloni R, Lancet D (2006) A probabilistic classifier for olfactory receptor pseudogenes. *BMC Bioinformatics* 7:393
58. Thierry-Mieg D, Thierry-Mieg J (2006) AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol* 7(Suppl 1):S12 1–S12 14
59. Young JM, Endicott RM, Parghi SS et al (2008) Extensive copy-number variation of the human olfactory receptor gene family. *Am J Hum Genet* 83:228–242
60. Menashe I, Abaffy T, Hasin Y et al (2007) Genetic elucidation of human hyperosmia to isovaleric acid. *PLoS Biol* 5:e284
61. Keller A, Zhuang H, Chi Q et al (2007) Genetic variation in a human odorant receptor alters odour perception. *Nature* 449:468–472
62. Kent WJ (2002) BLAT—the BLAST-like alignment tool. *Genome Res* 12:656–664
63. Camacho C, Coulouris G, Avagyan V et al (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421
64. Pearson WR, Wood T, Zhang Z et al (1997) Comparison of DNA sequences with protein sequences. *Genomics* 46:24–36
65. Go Y, Niimura Y (2008) Similar numbers but different repertoires of olfactory receptor genes in humans and chimpanzees. *Mol Biol Evol* 25:1897–1907
66. Olender T, Waszak SM, Viavant M et al (2012) Personal receptor repertoires: olfaction as a model. *BMC Genomics*. Aug 21;13:414. doi: 10.1186/1471-2164-13-414



<http://www.springer.com/978-1-62703-376-3>

Olfactory Receptors

Methods and Protocols

Crasto, C.J. (Ed.)

2013, XV, 259 p., Hardcover

ISBN: 978-1-62703-376-3

A product of Humana Press