The sense of smell: genomics of vertebrate odorant receptors

Janet M. Young and Barbara J. Trask*

Division of Human Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N., C3-168, PO Box 19024, Seattle, Washington 98109, USA

Received and Accepted February 20, 2002

Olfactory receptor (OR) proteins interact with odorant molecules in the nose, initiating a neuronal response that triggers the perception of a smell. The OR family is one of the largest known mammalian gene families, with around 900 genes in human and 1500 in mouse. After discounting pseudogenes, the functional repertoire in mouse is more than three times larger than that of human. OR genes encode G-protein-coupled receptors containing seven transmembrane domains. ORs are arranged in clusters of up to 100 genes dispersed in 40–100 genomic locations. Each neuron in the olfactory epithelium expresses only one allele of one OR gene. The mechanism of gene choice is still unknown, but must involve locus, gene, and allele selection. The gene family has expanded mainly by tandem duplications, many of which have occurred since the divergence of the rodent and primate lineages. Interchromosomal segmental duplications including OR genes have also occurred, but more commonly in the human than the mouse family. As a result, many human OR genes have several possible mouse orthologs, and vice versa. Sequence and copy number polymorphisms in OR genes have been described, which may account for interindividual differences in odorant detection thresholds.

The ability to detect smells serves essential and aesthetic functions in vertebrates. Many species rely heavily on their sense of smell to locate food, detect predators or other dangers, navigate, and communicate societal information. Olfaction plays a role in mate choice, mother–infant recognition and signaling between members of a group (1,2). Humans may not depend on olfaction for survival to the same extent as other species, but we nevertheless use olfaction extensively to gather information on our surroundings, from the student checking if her milk is still safe to drink, to the hunter using smell to help locate his prey, to the wine taster enjoying a vintage Bordeaux. Humans also communicate via odors and pheromones, both consciously (by applying artificial scents) and subconsciously. For example, olfaction mediates the curious synchronization of menstrual cycles of women living in close proximity (3). Aesthetically, olfaction is important for our enjoyment of food, our natural environment and our lives in general (4). Human olfactory abilities are not well characterized, but it has been estimated that we can detect, although perhaps not discriminate, thousands of different odorants.

AN ENORMOUS GENE FAMILY

How do the nose and brain detect such a large array of small volatile chemicals and process that information into the perception of a smell? Multiple odorant molecules, differing in shape and chemical properties, must be discriminated. This task is analogous to the recognition of diverse antigens by the immune system. Both the olfactory and immune systems accomplish molecular recognition and discrimination by expressing many different receptors in the responding tissue. In olfaction, as many as 1000 different types of receptors are collectively used to detect thousands of different odorant molecules. Each receptor type is expressed in a subset of neurons in the olfactory epithelium, from which neurons wire directly to the olfactory bulb in the brain (reviewed in 5). From here, signals are relayed to the olfactory cortex (6), where they result in the perception of smell.

The astute prediction that odorant receptors would comprise a large family of seven-transmembrane-domain G-protein-coupled receptors (GPCRs) expressed exclusively in the olfactory epithelium led to the cloning in 1991 of the first few olfactory receptor (OR) genes in rat (7). This finding opened the way for characterization of the OR gene family, which was predicted on the basis of genomic library screens to comprise 1000 members in mammals (8). Recently, exhaustive searches of the almost complete genome sequences of human and mouse have led to the identification of 906 human OR genes (9) and 1393 (10) or 1296 (11) mouse OR genes. These numbers extrapolate to a total of 950 human and 1500 mouse genes after missing sequences are accounted for (9,10). ORs are thus one of the largest known gene families in the human and mouse genomes, representing 3–5% of all genes (12). Several

*To whom correspondence should be addressed. Tel: + 1 206 667 1470; Fax: + 1 206 667 4023; Email: btrask@fhcrc.org
organism to detect a range of smells? It has been dif

How does this genomic collection of OR genes enable an

limited number of mouse genes (13–20) and a handful of

expression in the olfactory epithelium is available for only a

section). It has been shown that each of these genes functions as

an odorant receptor – most are denoted as such merely on the

basis of sequence similarity to known ORs. Evidence of

expression in the olfactory epithelium is available for only a

limited number of mouse genes (13–20) and a handful of

human genes (21–24). Even fewer OR genes have been

functionally expressed in heterologous cell types to confirm

that they respond to odorants and elicit a signaling cascade

(25–29). Expression has been reported in non-olfactory tissues,

principally the testis (30–35), but also taste tissues (36–39),

prostate (24), primordial germ cells (40), erythroid cells (41),

heart (42) and notochord (43). It is not clear whether the

ectopic expression of these genes has functional significance

(44), perhaps as spermatid chemoreceptors in chemotaxis

towards the egg (30), or even as a developmental addressing system (45). It is also possible that the transcripts observed in non-olfactory tissues are merely genomic contaminants in cDNA libraries and/or low-level illegitimate transcripts.

Olfactory receptor gene sequences have also been reported from many other vertebrate species, including phylogenetically diverse species as non-human primates (46), pig (47), dog (48), rat (7), dolphin (49), marmot (50), chicken (51,52), frog (53), coelacanth (49), catfish (54), zebrafish (55) mudpuppy (56), lamprey (57), medaka (58) and goldfish (59). Fish are estimated to have much smaller OR repertoires than mammals, with around 100 genes (54). More distantly related chemosensory receptor genes have been described in fruit flies (60–63) and nematode worms (64–66). Although all are 7-transmembrane domain GPCRs, the fly, worm, and vertebrate families appear to have arisen independently. Other mammalian families of more distantly related GPCRs, such as taste receptors (reviewed in 67) and pheromone receptors (reviewed in 68), are also involved in recognition of external chemical signals; these families also appear to have evolved independently of the OR family.

A startlingly high fraction of the human OR repertoire (at least 63%) has degenerated to pseudogenes, leaving only about 350 apparently functional OR genes (9,69,70). In contrast, only about 20% of mouse ORs are pseudogenes (10,11), giving mice over three times as many intact genes as humans. Furthermore, some apparently intact human OR genes lack motifs that are very highly conserved in intact mouse OR genes, suggesting that not all human genes encode functional OR proteins (10). These molecular differences may be the inevitable consequence of humans’ reduced reliance on smell relative to rodents (1).

However, it is not known whether their larger repertoire of functional genes enables mice to detect a wider range of odorants or to better distinguish closely related odorants than humans. When sequences are examined phylogenetically, intact human genes are found in most OR subfamilies (10,11). Thus, it is likely that humans can detect as wide a range of substances as mice, assuming that different OR subfamilies bind different odorant classes.

FROM GENES TO SMELLS

How does this genomic collection of OR genes enable an organism to detect a range of smells? It has been difficult to associate OR genes with their odorant ligands, but two approaches have shed some light on the question. In a technological tour de force, a number of OR–ligand pairings were identified by amplifying mRNA isolated from individual olfactory neurons responding to test odorants (71,72). Some groups have also succeeded in expressing OR genes in sensory neurons (28) or heterologous cell lines (25–27,29), or purifying neurons expressing the same OR (73), and measuring the response of those cells to a panel of odorants. Results from these approaches suggest that ORs bind odorants combinatorially: each OR can respond to multiple structurally related odorant molecules, albeit with varying response amplitudes, and a given odorant can elicit responses from multiple OR types (Fig. 1). Unfortunately, many attempts to express ORs in heterologous cells fail (74). A better understanding of how ORs are installed in membranes and transported to the cell surface may translate to improved in vitro assays of ligand–receptor relationships.

Olfactory binding proteins (OBPs) may also contribute to the combinatorial code. OBPs are a family of proteins secreted in the nasal mucosa that may help transport odorant molecules through the mucus (75). If OBPs remain associated with odorants during their interaction with ORs, their loose specificity (75) may add an extra layer of complexity to the problem of determining OR–ligand relationships.
FROM A CHEMICAL MIX TO A SPATIAL CODE

The basic organizing principles of the olfactory system enable the brain to interpret odorant-receptor interactions. Each sensory neuron in the nose expresses only one allele of one of the 1000 or so OR genes in the genome (14,71,76), although there are occasional exceptions to this rule (77). The mechanisms that restrict expression to a single gene probably evolved to prevent a neuron from being responsive to too many ligands to be informative to the organism. Monoallelic expression may simply be a by-product of the mechanism ensuring single gene expression, but it guarantees that allelic variants with different affinities are expressed in separate neurons, allowing for segregation of responses. Neurons expressing a given OR are found in one of four topographically defined zones of the olfactory epithelium, but appear to be randomly distributed within that zone (13,78). A xons of all olfactory neurons expressing a given OR converge on just one or two glomeruli in the olfactory bulb (79), and each glomerulus is probably innervated by neurons expressing only one OR gene. Response from neurons distributed across the olfactory epithelium is thus translated into a spatially organized neuronal response that can be relayed to the olfactory cortex (6,80,81).

The neurons of the olfactory epithelium are unusual in that they regenerate throughout the life of an individual; the entire epithelium turns over every few months (82,83). It is not yet known if the OR gene choice is made in the terminal differentiation step, or if the choice is fixed in the precursor cell, so that all of its progeny express the same receptor. Recent experiments with in vitro culturing systems suggest that the latter may be the case (84). Neuronal wiring to the correct glomerulus in the olfactory bulb is re-established during regeneration (85). More work needs to be done to understand how olfactory neurons converge during bulb formation in the fetus and when neurons regenerate later in life, but it is clear that expression of the OR gene itself is necessary for correct wiring (81,86).

GENE CLUSTERS AND SINGULAR EXPRESSION

How does a neuron select one allele of one of the 1000 genes to express? The olfactory and immune systems are uncannily similar in the singularity of receptor expression in responding cells, but the systems must rely on different mechanisms to achieve this. Singular expression is achieved in the immune system with efficient use of genomic resources: gene segments housed in a single, compact genomic locus are rearranged uniquely in each cell (87). In contrast, each OR appears to be expressed from a separate gene. The coding region is contained within a single exon (simplifying experimental and computational gene identification), but exons in the 5′ untranslated region may undergo alternate splicing (Fig. 2) (19,22,23,88–90). The OR genes are dispersed among loci that contain anywhere from 1 to over 100 genes in more than 40 chromosomal locations in the mouse genome (10) and over 100 locations in the human genome (9). The average spacing between neighboring OR coding regions is 21 kb (10), and non-OR genes are usually excluded from within OR clusters (9). Altogether, OR clusters occupy vast amounts of genomic territory: nearly 1% of the mammalian genome is devoted to olfaction (9).

The existence of multiple, widely dispersed gene clusters means that OR gene regulation must involve locus choice, gene choice within a cluster, and allele choice. Regulation may therefore involve both cis-acting and trans-acting processes, but none has been identified to date. OR genes in the same genomic cluster or phylogenetic clade are not necessarily expressed in the same zone in the olfactory epithelium as one another (15). Zone choice is therefore likely to involve cis binding of zone-specific transcription factors. Gene clustering may not be functionally important, but could simply be a consequence of the fact that tandem duplications have been the primary mechanism for OR family expansion (9,10). Transgenic experiments using two identical copies of an OR gene, distinguished by coexpression of different reporter molecules, show that the two copies are not expressed in the same neurons (i.e. singular expression holds even for identical genes) (91). Current ideas for how singular OR gene expression is achieved include recombination into an active locus (similar to the yeast mating-type switch), use of locus control regions, combinatorial transcription factor binding, stochastic activation by limiting amounts of transcription complexes, and epigenetic mechanisms. OR gene regulation has been reviewed recently (92), so we summarize results only briefly here.

![Genomic organization of olfactory receptor genes](https://academic.oup.com/hmg/article-abstract/11/10/1153/610765/2)

**Figure 2.** Genomic organization of olfactory receptor genes. Top: OR genes have a single main coding exon (black) and typically have several 5′ untranslated exons. Alternate splicing is seen in many genes. Middle: OR genes are clustered in the genome in groups of 1 to over 100 genes (green arrows) and pseudogenes (red arrows) in both transcriptional orientations. Bottom: OR clusters are dispersed around the genome in more than 40 (mouse) or over 100 (human) locations.
Experiments using transgenic mice suggest that OR gene expression can be correctly driven by sequences local to the gene (93), but seemingly irreconcilable results with different constructs suggest the presence of both positive and negative regulators in the region surrounding OR genes, some operating at long range (91,94). Neither in vivo experiments nor genomic analyses have convincingly implicated particular regulatory motifs. Comparison of orthologous human and mouse genomic sequences or sequences upstream of paralogous genes of one species - a strategy that has been used successfully to find conserved regulatory sequences of other genes (95) - has identified putative regulatory features upstream of some OR genes (55,88) but not others (19,20,89). A plethora of transcription binding factor sites are found, but no patterns have emerged that might explain how a gene is chosen among the many genomic possibilities. This paucity of obvious regulatory clues stands in contrast to the V1R genes, another family of GPCRs that is involved in pheromone detection (96). V1Rs also exhibit monoallelic (and probably monogenic) expression in vomeronasal neurons (97), and large conserved regions (> 500 bp) found upstream of many mouse V1R genes are candidate regulatory regions (98). Improved computational tools to detect subtle regulatory signals and in vitro experimental systems capable of systematically assaying for functional elements in expression constructs are badly needed to understand OR regulation.

A DYNAMIC GENE FAMILY

The OR gene family is evolving rapidly. As is the case for major histocompatibility alleles, the pool of allelic variants of each gene in the human population is likely to be large (99,100). Some OR genes have become pseudogenes very recently during primate speciation (101,102) and even during the spread of the human population (99,103). As a result, different people could have different complements of functional OR genes. Patterns of nucleotide and amino acid changes (54,58,104,105), as well as population studies of polymorphisms (106), show that some regions of OR genes are subject to positive selection (i.e. after gene duplication, change in protein sequence is advantageous, giving rise to a more diverse receptor repertoire).

Most genomic locations that now contain OR genes did so when primate and rodent lineages diverged (i.e. orthologous mouse–human locations can be identified for most extant clusters) (10), but within those locations, there have been many local duplications of genes (and flanking regions) in both lineages since divergence (10,19,89,107,108) (Fig. 3). Many human OR genes therefore do not have a single clear ortholog in the mouse, and vice versa. In fact, local duplications, likely mediated by unequal recombination, are the major force creating new OR genes (9,10) (Fig. 3). The patterns of change seen between neighboring OR gene duplicates suggests that gene conversion can be an additional force shaping gene sequence (102).

Other OR genes have undergone segmental duplications to distant chromosomal locations. Ancient duplications involving entire OR genomic clusters were probably responsible for the dispersal of OR genes around the genome (9). More recent non-local duplications have involved large blocks of tens to hundreds of kilobases of sequence, often near the telomeres and centromeres of chromosomes (10,109-111). Interchromosomal OR duplications are much more common in human than mouse (10). A subset of subtelomeric OR duplications are so recent that the copy number and locations of the duplicated block are polymorphic (100,110). Some of these OR genes appear to be functional, and multiple copies are expressed (23). Such copy-number polymorphism is easily detected by fluorescence in situ hybridization (110). Copy-number polymorphism of recent local gene duplicates may also exist, but will be more difficult to detect and analyze - sequence differences between paralogs have already been mistaken for polymorphisms (112).

Not all interchromosomal OR duplications involve subtelomeric regions. For example, the OR7E subfamily of human pseudogenes has recently duplicated to at least 35 genomic locations (9,109). The molecular mechanism for these large segmental duplications is unknown, but is unlikely to involve unequal crossing-over (113).

Figure 3. The OR gene family has expanded primarily by local duplications. This point is illustrated by a phylogenetic tree of all full-length mouse and human OR protein sequences and a chemokine outgroup, with genes from a selected pair of syntenic genomic locations highlighted. Mouse sequences are colored red and human sequences blue. A II sequences mapping to mouse chromosome 2E1 and the syntenic human location, chromosome 11p11-11q12, are colored darker red and blue, respectively. Sequences from these pairs of locations cluster together on the tree, confirming that orthologous clusters have been identified. Clusters of like-colored branches illustrate that (i) local duplications are a major force shaping the OR gene family (e.g. human 11p11-11q12 sequences are found in only two major clades) and (ii) many of these local duplications have occurred since rodent-primate divergence. Adapted from Young et al. (10).
OR EVOLUTION AND FUNCTIONAL DIVERSITY

Recent expansion of the OR gene family may have enabled mammals to detect and discriminate an increasing number of odorant molecules. Positive selection to increase sequence diversity between closely related receptors may increase discriminatory powers, since even a few changes in amino-acid sequence can be sufficient to change odorant specificity (27). The regulatory regime that selects a single gene for expression guarantees that variant alleles or recently duplicated genes are expressed in different cells, and consequently, mutations that confer an advantageous new function can be selected for. It is not clear how different two genes or alleles must be in order for their neurons to be wired to different glomeruli, a change which is presumably required for a shift in perception. However, sequence and copy-number differences in individual OR gene repertoires are likely to translate into altered sensitivity to and discrimination of a number of odorants. Such phenotypic differences have been reported both in mice, with genetic differences at a locus on chromosome 6 causing some strains to be anosmic for isovaleric acid (114), and in humans, where heritable differences in the ability to detect several odorants have been described (115,116), but not yet genetically mapped.

The plasticity of the OR subgenome can extend beyond olfactory function. Members of the OR7E subfamily of human genes have duplicated after becoming pseudogenes (9,111), making it unlikely that olfactory-based selective pressure has favored such duplications. These highly similar sequence blocks of OR genes in distant genomic locations can in fact be disadvantageous, since they can mediate genomic rearrangements. For example, two widely separated copies of a block containing OR7E pseudogenes on human chromosome 8 occasionally recombine to cause chromosomal abnormalities with associated mental handicap (117).

FUTURE DIRECTIONS

The size and complexity of the OR gene family leave us with many challenges. Sophisticated experimental and computational tools are required to understand the mechanism of singular expression of OR genes and to link ORs with their odorant ligands. Study of OR gene diversity, polymorphism and evolution, especially of recent tandem and segmental duplications, will translate to a better understanding of olfactory biology, phenotypic variation, and the processes shaping mammalian genomes.

ACKNOWLEDGEMENTS

The authors thank Tera Newman, Bob Lane, Mike Schlador and Eleanor Williams for comments on the manuscript and providing ideas and material for the figures, other members of the Trask lab for helpful discussions, and NIH Grant R01 DC04209 for financial support.

OR databases and nomenclature

OR gene sequences are available from several partially overlapping databases - GenBank (118), HORDE (9) (http://bioinformatics.weizmann.ac.il/HORDE/), ORDB (119) (http://senselab.med.yale.edu/senselab/ordb/) and a mouse OR database that also contains orthology information (10) (http://www.fhcrc.org/labs/trask/OR/). Nomenclature of OR genes is confusing - systematic naming schemes have been proposed for human (120) and mouse (11) OR genes, but have not been widely adopted. Lack of coordination on nomenclature has lead to the existence of several names for the same gene, and even some examples of different genes with the same name. The difficulties of distinguishing allelic variants from closely related paralogs (100,112) will add to this confusion.

REFERENCES


