Novel human vomeronasal receptor-like genes reveal species-specific families

RODRIGUEZ, Ivan, Mombaerts, Peter

DOI: 10.1016/S0960-9822(02)00909-0
Novel human vomeronasal receptor-like genes reveal species-specific families

Ivan Rodriguez¹,² and Peter Mombaerts¹

Mammals have a powerful sense of smell. In rodents, three unrelated superfamilies encoding seven-transmembrane domain proteins are expressed in sensory epithelia within the nasal cavity and encode chemosensory receptors: odorant receptors [1], type 1-vomeronasal receptors (V1rs) [2], and type 2-vomeronasal receptors (V2rs) [3-5]. In rodents, V1r and V2r genes are expressed by vomeronasal sensory neurons, a cell population known to be responsive to pheromones [6,7]. V1r genes, but not V2r genes, have intronless coding regions, facilitating their identification from genomic sequence. Previously we reported reiterative similarity-based searches for mouse V1r genes, taking advantage of recent significant progress in the sequencing of the mouse genome. We thus discovered 8 novel, highly divergent V1r families, grouped in a total of 12 families that comprise at least 137 V1r receptor genes with an intact open reading frame (ORF), in addition to numerous pseudogenes [8]. We and others have reported a single human V1R-like gene (hV1RL1) with an intact ORF from genomic DNA [9,10], and we showed that its spliced mRNA is present in human olfactory mucosa [9]. All other reported human V1R-like sequences are pseudogenes [11,12], suggesting a decline in the functional compartment of the V1R superfamily during evolution. Other than from mouse, rat, goat and human, no other mammalian V1Rs have been described.

Here, we provide a first global draft of the human V1R repertoire, based on genomic database mining. The tblastn algorithm was used to compare amino acid query sequences to the DNA databases of the National Center for Biotechnology Information database resources (nr, hhtgs, est divisions of GenBank) and the Celera Discovery System database, with a non-stringent expectation cutoff value of 1e-5. The queries included multiple sequences from each of the 12 mouse V1r families [8]. Sequences with ORFs encoding polypeptides shorter than 285 amino acids, or missing the first or the last putative seven-transmembrane domain, were not considered as full-length, uninterrupted ORFs. Sequences showing more than 98% identity at the nucleotide or amino acid level were considered as identical. These stringent criteria were motivated by our desire to exclude pseudogenic sequences, but may have excluded a few genuine human V1R genes. A total of ~200 human V1R-like sequences were identified, among which 5 had an intact ORF (termed hV1RL1-5). V1R-like pseudogenes were found on all human chromosomes except chromosome 20, with more than 40 on chromosome 19.

A few members of the large superfamily of rodent V1r genes could have kept an intact ORF in humans merely by chance. To address this question we identified 14 amino acid residues that are conserved in more than 95% of mouse V1r genes [8]. Assuming that many of these residues represent structurally relevant amino acids in the V1r superfamily, they should be conserved in functional V1r members from other mammalian species. In accordance with this expectation, hV1RL1-5 have respectively 14, 12, 12, 12 and 6 of these 14 residues conserved (Figure 1). In addition, a potential N-linked glycosylation site (NXS/T) in extracellular loop 2 is conserved in more than 98% of mouse V1r genes [8]; again, this site is conserved in hV1RL1-4. We performed a Fisher’s exact test comparing the conservation of these residues between hV1RL1-4 and 23 hV1R pseudogenes, using 7 conserved internal amino acids together with the conserved glycosylation site; the one-tailed p-value is 0.000855. To address the possibility of a recent triplication of a founder pseudogene having led to the generation of the three related hV1RL2-4 genes, we compared their nucleotide similarities, considering separately the coding sequences, and the immediate 5’ and 3’ flanking non-coding sequences. Identities range from 61 to 64% for coding sequences, from 17 to 24% for the 500 bp

Figure 1. Alignment of the deduced amino acid sequences of the 5 human V1R-like genes containing an uninterrupted ORF. In red are the 14 conserved residues found in mouse V1rs. Green boxes show the N-linked glycosylation sites. The cV1RL1 sequence was obtained by amplifying chimpanzee genomic DNA with degenerate primers specific for hV1RL1. The dV1RL1 sequence was extracted from a dog (Canis familiaris) genomic draft sequence (GenBank AC090002). hV1RL1 has previously been reported by us as GenBank AF255342 [9]; hV1RL2, hV1RL4 and hV1RL5 were extracted from draft sequences GenBank AC010467, AL092070 and AC390728 respectively. hV1RL3 was previously identified (GenBank AF336873), but shows discrepancy with the Celera database, in which it is a pseudogene. We therefore subcloned and sequenced hV1RL3, and confirmed the GenBank AF336873 sequence with an intact ORF. GenBank accession numbers for hV1RL2,4,5, cV1RL1, dV1RL1 are AF370359, AY114733, AY114735, AY114734 respectively.
current biology

Clades on phylogenetic trees, and naturally fall into clearly distinct extremely isolated: first, they families are phylogenetically mouse V1r superfamily [8]. The its way to gene death.

Massively decayed repertoire on coincidental remnants of a these human genes are all function of these receptors in these evolutionary conservations pseudogene origin. Together, thus arguing against a downstream of the stop codon, 25 to 34% for the 500 bp upstream of the start codon, and a second cutoff at 15% corresponding to the minimal identity between V1rs. 

Figure 2.

(A) Unrooted tree of V1r and V1RL receptors. The 5 human (hV1RL1-5), the dog (dV1RL1) the chimpanzee (cV1RL1) deduced amino acid sequences were compared to the complete mouse V1r repertoire containing 12 families termed V1ra to V1rl [8]. Amino acid sequences were aligned using ClustalX. The tree was generated by using the neighbor-joining method with 1000 bootstrap replications. (B) Ranges of amino acid sequence similarities among and within human V1R families. In gray shading, the human sequences are joining method with 1000 bootstrap replications. (B) Ranges of amino acid sequence similarities among and within human V1R families. In gray shading, the human sequences are

upstream of the start codon, and 25 to 34% for the 500 bp downstream of the stop codon, thus arguing against a pseudogene origin. Together, these evolutionary conservations are consistent with a biological function of these receptors in humans. We deem it unlikely that these human genes are all coincidental remnants of a massively decayed repertoire on its way to gene death.

We distinguish 12 families in the mouse V1r superfamily [8]. The families are phylogenetically extremely isolated: first, they naturally fall into clearly distinct clades on phylogenetic trees, and second, interfamily amino acid identities are typically below 40%, while intrafamily identities are above 40%. Interestingly, HVR1Ls do not fit into any mouse V1r family: they form separate clades when compared to mouse V1rs (Figure 2A), and they do not share the minimum of 40% of their residues with all members of a given mouse family (Figure 2B). However, hV1RL2-4 share moderate interfamily identities with mouse families V1re, f, g and l (Figure 2B). Similar marginal interfamily identities were encountered with mouse families V1ra/b, V1re/f and V1rh/i [8]. If only a few functional V1Rs in a given species were conserved during evolution, one would expect them not to be kept at random, but members of specific families to be selected. We observed such a bias: hV1RL2-4, which are closely related in sequence, form their own separate clade when compared to mouse V1rs (Figure 2A). hV1RL5 appears very distant both from mouse and human V1r sequences, and may thus represent a genuine founder of a new family; however, due to the high number of mismatches with the previously described 14 conserved residues, this gene may be a pseudogene that kept an intact ORF by chance.

The intriguing possibility that mammalian species may differ significantly in the composition of their V1R repertoires was further explored by searching for V1R sequences of non-rodent species. We identified intact V1R ORFs in dog and chimpanzee genomic sequences, which fit into the clade formed by hV1RL1 (Figure 2A), indicating that the V1R repertoires of different species may partially overlap. As these receptors are candidates for the regulation of reproductive and other social behaviors, rapid evolutionary divergence in the V1R repertoire may underlie speciation by contributing to reproductive isolation. Such a view would be in line with the phylogenetic analyses of mouse V1rs, which point to a history of recent expansion of the V1R repertoire [8,13].

It should not be overlooked that despite several indirect lines of evidence, there is no proof that any of these receptors, including the mouse V1rs, is a pheromone receptor. In rodents, V1rs are expressed in the vomeronasal organ, an olfactory structure that is probably non-functional in humans; however, detection of certain pheromone cues in mammals is independent of the vomeronasal organ [14]. Expression of hV1LR1 in the human main olfactory mucosa [9] shows that it likely represents a chemosensory receptor. But this or the other human receptors could have been co-opted to fulfill chemosensory functions distinct
from detecting pheromones, as defined in the classical sense.

References

The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.
Department of Zoology and Animal Biology, University of Geneva, CH-1211 Geneva 4, Switzerland. E-mail: Ivan.Rodriguez@zoo.unige.ch, peter@rockefeller.edu