Sequence analysis of mouse vomeronasal receptor gene clusters reveals common promoter motifs and a history of recent expansion

Robert P. Lane*†‡, Tyler Cutforth§, Richard Axel§, Leroy Hood*¶, and Barbara J. Trask*†

*Department of Molecular Biotechnology, University of Washington, Seattle, WA 98195; and §Department of Biochemistry and Molecular Biophysics, and Howard Hughes Medical Institute, College of Physicians and Surgeons, Columbia University, New York, NY 10032

Contributed by Richard Axel, November 14, 2001

We have analyzed the organization and sequence of 73 V1R genes encoding putative pheromone receptors to identify regulatory features and characterize the evolutionary history of the V1R family. The 73 V1Rs arose from seven ancestral genes around the time of mouse–rat speciation through large local duplications, and this expansion may contribute to speciation events. Orthologous V1R genes appear to have been lost during primate evolution. Exceptional noncoding homology is observed across four V1R subfamilies at one cluster and thus may be important for locus-specific transcriptional regulation.

In most mammals, olfactory sensory perception is mediated by two anatomically and functionally distinct sensory organs: the main olfactory epithelium (MOE) and the vomeronasal organ (VNO). The MOE detects a vast repertoire of odors that provide information about the environment at large. The VNO is thought to recognize a more restricted array of odors, including pheromones, that provide information about the social and reproductive status of other individuals and elicit innate behavior and neuroendocrine responses (1–4). Although pheromonal substances remain largely uncharacterized in mammals, distinct social and sexual behaviors exhibited by different species are likely to be reflected in differences in the repertoire of pheromone receptors.

The odorant receptors expressed in the two sensory organs are encoded by distinct gene families. In mice, receptors in the MOE are encoded by a family of about 1,000 genes (5–7). The VNO receptors are encoded by at least two distinct gene families: ≈100–200 V1R genes are expressed in cells in the apical compartment of the VNO (8, 9) and ≈100 V2R genes are expressed in the basal domain (10–12). Each of the three gene families encodes highly divergent G protein-coupled receptors with seven transmembrane (TM) domains. Individual neurons in both the MOE and VNO express one allele of a single receptor gene such that the function of the sensory cell is defined by the receptor gene that is transcribed (13, 14). The mechanisms that assure the restricted expression of a single receptor in both MOE and VNO neurons remain elusive.

We have conducted a detailed genomics analysis of genes in two major mouse V1R loci on chromosome 6 and a third V1R locus on chromosome 13 to study the evolution and regulation of V1R genes. The three loci contain at least 73 V1R genes, including 43 with ORFs. A large region of putative regulatory homology is found upstream of V1Rs at one locus, but is not present at the other loci. These clusters arose from seven ancestral V1Rs through duplications of ≈5- to 10-kb DNA segments around the time when mice and rats diverged. This expansion could meet adaptive requirements during speciation. Examination of available human genomic sequence indicates that the V1R genes corresponding to these mouse loci have dispersed or been deleted during primate evolution.

Methods

Clone Isolation and Characterization. Bacterial artificial chromosome (BAC) clones were isolated from a 3-fold redundant BAC library derived from mouse (strain 129 SVJ) embryonic stem cells (Genome Systems, St. Louis). PCR primers designed from the 3′ untranslated regions (UTRs) of *V1Rb1* (VN2) and *V1Ra1* (*VN12/mV1R1*) receptor sequences (13, 14) were used to screen BAC library pools. Three positive clones were identified (BACs 27e23, 83g9, and 112m5). All clones were mapped to mouse metaphase chromosomes by using standard procedures for fluorescence *in situ* hybridization (15).

Shotgun Library Construction and Sequencing. BAC 27e23 was sequenced by the shotgun method. Including PCR-directed finishing, 2,590 reads (7.8× redundancy over the 198,517-bp BAC insert) were assembled into contiguous sequence by using PHRED/PHRAP (16, 17) and CONSED (18) assembly software. The estimated error rate is 1 in 2 × 10^4 nt (PHRED/PHRAP). The annotated sequence is in GenBank with accession no. AF129005.

Nomenclature. For the mouse 6D cluster, we adopted nomenclature used by Del Punta *et al.* (19). In some figures, these names are abbreviated (e.g., *V1ra1* to *a1*). An additional *V1Ra1*-subfamily member, *Y12724*, is named here according to its GenBank accession number. The V1R genes identified on the GAx5J887W5NCT, GAx5J887W5BDS, and GAx5J8B7W52BC Celera scaffolds are named with a prefix to specify the scaffold (NC, BD, or BC) followed by a number to specify relative position within the scaffold. V1R genes identified in BAC RP23–9O16 are named with the prefix 13, to reflect its annotated location on mouse chromosome 13, followed by a number to specify relative position within the BAC.

5′ Rapid Amplification of cDNA Ends (RACE) PCR. Poly(A)$^+$ mRNA was prepared from 10 young-adult (<4 wk) mixed-sex C57BL/6 VNOs by using the Quick Prep Micro RNA Extraction Kit (Amersham Pharmacia). cDNA synthesis and 5′ RACE was performed by using the Marathon cDNA Amplification Kit (CLONTECH).

Genomic Analysis Tools. Repeat content was determined by the REPEATMASKER algorithm (http://ftp.genome.washington.edu). Sequence analyses were done by using MEGALIGN (DNAstar,

Abbreviations: MOE, main olfactory epithelium; VNO, vomeronasal organ; TM, transmembrane; BAC, bacterial artificial chromosome; UTR, untranslated region; RACE, rapid amplification of cDNA ends; L1, Line 1.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AF129005).

†Present address: Human Biology Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109.

‡To whom reprint requests should be addressed. E-mail: rlane@fhcrc.org.

¶Present address: Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105.

www.pnas.org/cgi/doi/10.1073/pnas.012608399

Madison, WI), PIPMAKER (20), MATINSPECTOR/TRANSFAC (21), and TSSG/TSSW/NNPP promoter prediction algorithms (http://searchlauncher.bcm.tmc.edu), Ka/Ks substitution ratios were calculated by using the DIVERGE program (GCG). Human V1R searches were performed on the April 2001 assembly (http://www.genome.ucsc.edu) by using TBLASTN.

Results

V1R Phylogeny Suggests Recent Expansion Within Old Clusters. To identify genomic features that would be informative about V1R regulation and evolution, we analyzed the sequence of a major V1R cluster on mouse chromosome 6. Three BAC clones were isolated that contain the previously characterized *V1Rb1* and *V1Ra1* genes (14, 19, 22). These BAC clones were localized to mouse chromosome 6D by fluorescence *in situ* hybridization (data not shown). High redundancy (≈8×) sequencing of the largest clone, BAC 27e23, resulted in ≈198.5 kb of contiguous genomic sequence (Fig. 5, which is published as supporting information on the PNAS web site, www.pnas.org). Additional draft sequences available in public and Celera databases were assembled into an ≈800-kb contig encompassing this finished sequence and extending ≈500 kb upstream and ≈100 kb downstream of the *V1Rb1* and *V1Ra9* genes, respectively (Fig. 1*A*).

This locus contains 23 V1R genes, including 16 with ORFs. With the exception of *V1Ra9*, all of these genes have been involved in recent duplications. The duplicated gene-containing blocks are ≈3–19 kb in size and exhibit 87–99% identity in the coding region and 78–97% identity in the duplicated DNA outside the coding exons. These analyses suggest that a not too distant rodent ancestor likely possessed four V1R ancestral genes at this locus. Subsequent expansion of three ancestral genes resulted in six *V1Ra1*-like ORFs, seven *V1Rb1*-like ORFs, and two *V1Ra7*-like ORFs.

V1R gene block duplications are a common feature of mouse V1R loci; two additional V1R clusters identified in mouse draft sequence exhibit a similar history of recent expansion. The draft sequence of a BAC annotated to mouse chromosome 13 contains nine V1R genes (five ORFs). Three Celera genomic scaffolds encompassing 1.35 Mb and 42 V1R genes (23 ORFs) are mapped by Celera ≈30–40 Mb centromeric to the 6D cluster. A molecular tree of the 44 ORFs from all three loci is shown in Fig. 2. V1R genes group monophyletically according to location: V1R coding regions are more than 65% diverged between loci (ref. 19; Fig. 2). The gene-containing blocks in these two additional loci, like the 6D locus, have duplicated recently (Fig. 1*B* and not shown). The noncoding portions of the duplicated blocks within these two loci are ≈68–77% identical. All 42 V1Rs at the second chromosome 6 locus and the nine V1Rs at the chromosome 13 locus appear to have arisen recently from one and two ancestral genes, respectively (Fig. 1*B* and not shown). Therefore, recent duplications have occurred within, but not between, loci.

V1R Loci Are Densely Populated with Repetitive Elements. The genomic regions surrounding the duplicated blocks at the mouse 6D V1R locus are densely populated with repetitive elements. More than 90% of the interblock sequence is classified as repeats by the REPEATMASKER algorithm (Fig. 1*A*). Simple repeats represent >1% of the total sequence and include long dinucleotide tracks between duplicated blocks. The Line-1 (L1) repeat family collectively comprises ≈250 kb (42%) of the locus, and most of these repeats are found between duplicated blocks.

The combined repeat content for the 2.1 Mb of the three V1R loci is 58%, a higher density than found in random mouse genomic sequence with equivalent GC content (42%; ref. 23). Approximately 70% of the repeat content is the L1 family, and ≈50% of the L1 content belongs to either the L1_MM and Lx subfamilies. The mouse 6D locus is especially abundant in L1_MM repeats (≈31% of its L1 content), whereas the other two loci are especially abundant in Lx repeats (≈40% of their L1 content). Because Lx repeats are ancestral to L1_MM repeats (24), the 6D locus has experienced more recent repeat activity than the other two loci.

The blocks themselves contain few repeats and are not flanked by common repeats (Fig. 1). L1_MM repeats within blocks are not in common relative positions, indicating that these duplications occurred before the L1_MM insertions. In contrast, several Lx elements were duplicated as part of larger blocks, indicating that these duplications occurred after Lx activation. These data suggest that many of the V1R duplications took place between the periods of Lx and L1_MM activity.

Duplication Timing Suggests Rapid V1R Expansion. The molecular tree (Fig. 2) suggests that expansion of the V1R repertoire at these loci occurred in bursts: within each clade, the terminal branches are of similar length and coalesce at approximately the same positions. We determined substitution rates of duplicated blocks to further investigate the duplication timing (Fig. 3). Based on their similarity, the duplications at the mouse 6D locus appear to have occurred more recently on average than those at the other two loci. Most of the blocks at the 6D locus have diverged 16–22%, compared with 23–32% at the other two loci. Assuming a constant mutation rate of 5×10^{-9} nt/yr for neutral sequence in rodent evolution (25), the majority of the expansion at the 6D locus began ≈22 million years ago and at the other two loci, 32 million years ago.

The narrow range of noncoding identities between blocks suggests that the bulk of the duplication events leading to the expansion of these subfamilies occurred over narrow periods of evolutionary time. Indeed, frequency distributions of pairwise divergences are significantly nonlinear (not shown), indicating that duplications have not occurred at a uniform rate since they began 20–30 million years ago. Rather, a burst of duplications appears to have occurred at these V1R loci.

Selective Pressures on V1R Coding Sequences. A comparison of synonymous versus nonsynonymous substitutions in coding regions shows that the duplicated V1Rs have been subject to selection both for and against change in specific domains (Fig. 6, which is published as supporting information on the PNAS web site). The average Ka/Ks ratio for all pairs of V1R ORFs at each of the three loci is ≈0.51, suggesting that these genes are generally under negative selection. The third TM domain is under particularly strong negative selection (average Ka/Ks, 0.29). The fourth TM domain, a region thought to interact with odorants and be under positive Darwinian selection in OR genes (26), exhibits a high average Ka/Ks (0.89), with nine pairs of V1R sequences having ratios >3.0. Selection for amino acid change in this region may have diversified the odorant-binding functions of some duplicated V1Rs. Unexpectedly, the loop between the first two TM domains shows Ka/Ks >1.0 overall and >3.0 for 31 V1R pairs. Positive selection for change in this region of the protein is surprising, because intracellular loops are not likely to play a role in binding odorants.

V1R Homology in the Human Genome. A search of the human genome identified putative V1R homologs on 17 human chromosomes and 42 distinct locations. The 53 human sequences identified had TBLASTN E-value scores of 1×10^{-51} to 1×10^{-5}, suggesting legitimate V1R homology. Included among these sequences were the eight V1R-like homologs previously reported (27, 28).

Only three putative clusters of human V1R homologs were identified. These are on human chromosomes 1 (270.7 Mb from pter), 7 (62.8 Mb), and 19 (65.7 Mb), where three V1Rs were found within 100 kb of each other. We found no evidence of a human V1R cluster orthologous to the mouse 6D V1R locus.

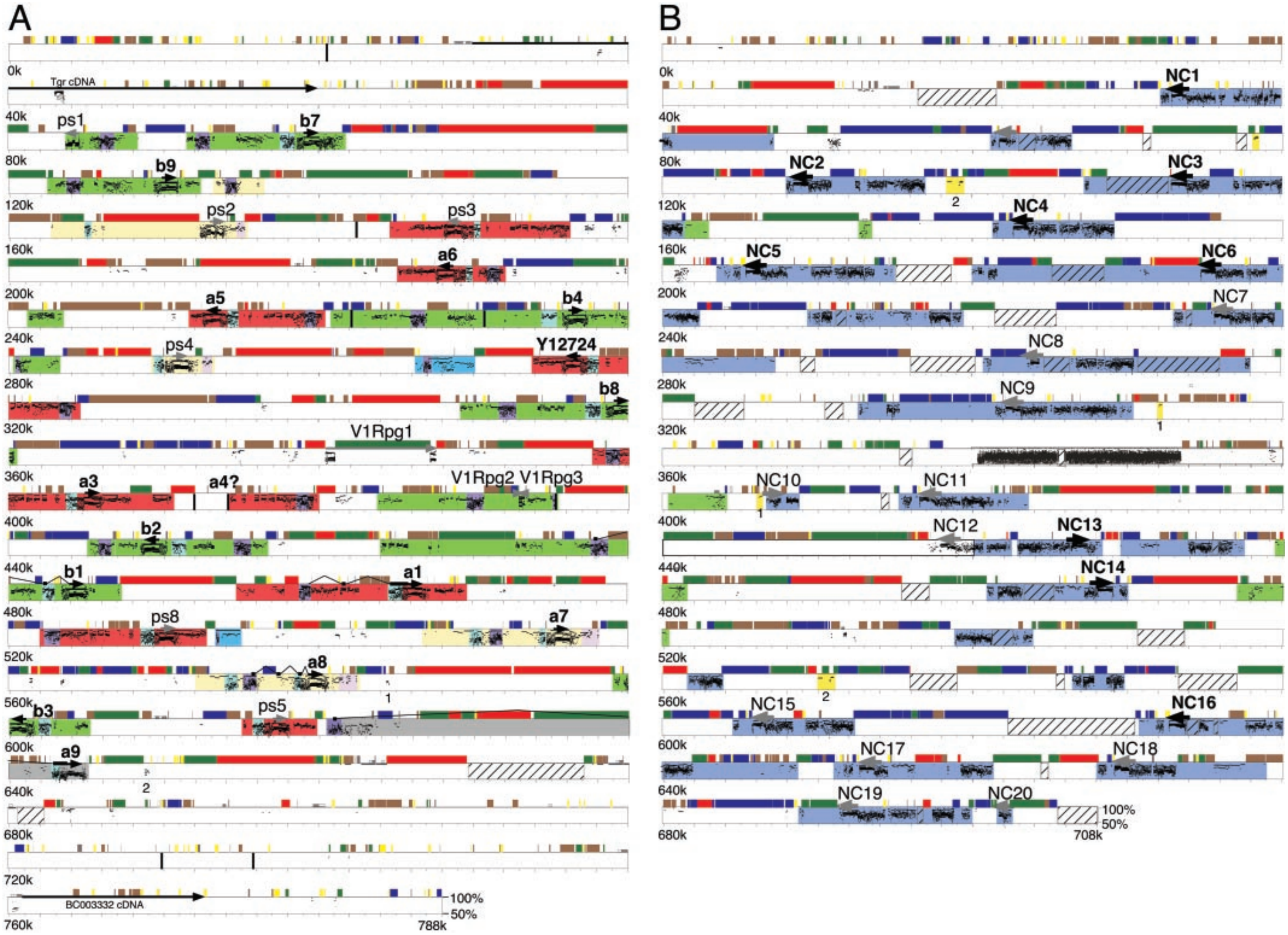

Fig. 1. PIPMAKER plots illustrating gene-block duplications and putative regulatory homologies at two mouse V1R loci. (*A*) The 788-kb sequence encompassing the mouse chromosome 6D V1R cluster is masked for repeats and compared with itself in both orientations. Homology is plotted according to position (horizontal axes) and percent identity (vertical axes, 50–100%). The positions of the V1R coding sequences with ORFs are indicated by black arrows above the plots (pseudogenes are indicated by gray arrows); the arrow direction indicates relative orientation. In contrast to a previous analysis of gene content and order by using PCR (19), we find *V1ra1* and a novel pseudogene *V1Rps8* located between *V1Rb1* and *V1Ra7*. The 5′ UTRs, where determined by RACE-PCR, are indicated by medium-thick lines (exons) and thin lines (introns). The locus is bounded by two multiexon genes: the *Tgr* cDNA at 30–80 kb and the BC003332 cDNA between 761 and 773 kb (black arrows). Celera contigs are ordered in the scaffold by paired-end sequence assembly. Vertical bars within the plot indicate positions of unresolved gaps in the assembled sequence; hatched boxes indicate gaps of estimated size. Duplicated blocks are color-coded: red blocks are duplications of the A subfamily, green blocks are duplications of the B subfamily, and yellow blocks are duplications of the a7/a8 subfamily. The a9 gene block (not duplicated) is shaded gray. Repeat content is summarized along the top axis (light gray, low complexity; dark gray, simple repeats; brown, LTR content; yellow, all SINE repeats; red, L1_MM L1 repeats; blue, Lx L1 repeats; and green, all other Line repeats). Putative regulatory regions are shown with color-coded shading: purple patches are the common promoter homology, light blue patches are the common 5′ UTR homology, pink patches are the locus control region homology (see text). Regions marked with medium blue shading at 308 kb and 534 kb are homologous to each other and contain the common promoter homology (purple portion), but do not appear to be associated with V1R genes or any larger duplicated blocks. Unshaded PIPMAKER signals are local repetitive sequence. The portion of finished sequence contributed by us (RP23–27e23) lies between 467 kb and 666 kb. (*B*) The 708-kb scaffold encompassing a portion of a second mouse V1R cluster mapped by Celera to chromosome 6.56 (56 Mb from pter on chromosome 6) is similarly masked for repeats and compared with itself. The 20 V1R gene-block duplications are shaded blue. Regions shaded yellow share homology to the 6D locus. These regions are homologous to positions at ≈585 kb and ≈649 kb on the 6D map in *A* (regions marked 1 and 2). Regions shaded green are putative regulatory regions that share homology to each other, as well as other regions in the genome annotated for putative regulatory function (e.g., T cell receptor Vα gene promoter regions and DNase1 hypersensitivity sites). The region between 380.5 kb and 393.5 kb is a microsatellite-like repetitive sequence. All other features and labels are as described for *A*. See Fig. 8, which is published as supporting information on the PNAS web site, for an enlarged version of this figure.

The *SPR* and *EGR4* genes flank the mouse 6D cluster (19), and human orthologs to *SPR* and *EGR4* are ≈325 kb apart on 2p13. BLASTX searches of the sequences of two BACs (AC012366.6 and AC010913.6) spanning this region of 2p13 identify no V1R homology. Therefore, either the mouse 6D V1R cluster inserted between *SPR* and *EGR4* since the mouse–human split or the ancestral V1R cluster has dispersed or been deleted in primate evolution.

The four human V1R-like sequences encoding the longest ORFs are on chromosome 19 and are most similar to the mouse 13.6 gene. These four homologs include the previously identified *V1RL1* ORF (28) at 71.2 Mb, plus a cluster of three V1R-like ORFs ≈5 Mb away. The first two lack a TM7 domain and are probably nonfunctional. The third V1R homolog may be a functional gene with a novel structure; it lacks at ATG start codon at the expected position and encodes an extended N terminus of ≈80 aa with low homology to tryptophan decarboxylase.

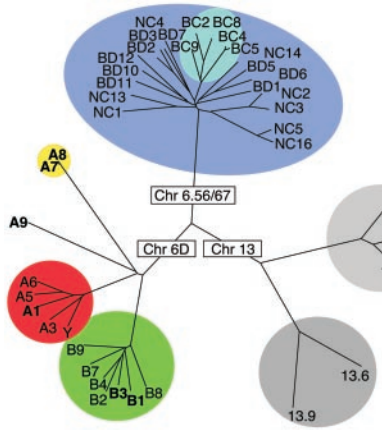

Fig. 2. Molecular tree illustrates local gene expansion. Unrooted PAUP (Sinauer Associates, Sunderland, MA) nucleotide distance tree of 44 V1R genes with ORFs from three chromosomal locations. The tree partitions into three monophyletic clades: one that contains all 16 V1R ORFs from the 6D locus, another that contains all 23 V1R ORFs from second region on chromosome 6 (6.56/6.67), and the third that contains all 5 V1R ORFs from the chromosome 13 BAC. Seven subfamilies (>25% nucleotide divergence in coding sequence) are color-coded: A subfamily, red; B subfamily, green; a7-a8 subfamily, yellow; a9 gene, unshaded; NC/BD/BC, blue (with BC subclade in light blue to reflect the distinct map location of this gene set); and the two chromosome 13 subfamilies, light and dark gray. Nomenclature is described in *Methods*.

Gene Structure: Shared Putative Promoters. The complete nucleotide sequence for these V1R genes, along with a transcriptional analysis, enables us to determine the precise intron/exon structure of V1R genes and may unveil common regulatory motifs that control receptor gene expression. We therefore have generated RACE–PCR products for four V1R genes to identify the transcription start site and putative upstream promoter regions. These cDNAs, when mapped onto genomic sequence, indicate a single coding exon and one or more 5′ exons (Fig. 1*A*). This gene structure is similar to that of OR genes expressed in the MOE (29–33).

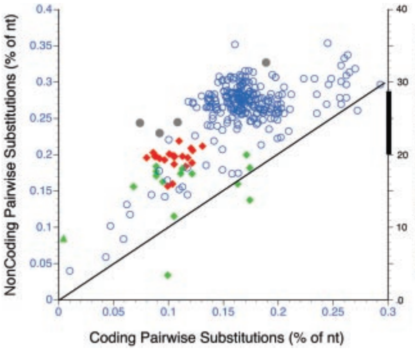

Fig. 3. Pairwise sequence divergence for noncoding and coding regions of duplicated V1R gene blocks. The duplicated blocks were compared within six of the seven V1R subfamilies (a9 is excluded, because it has not duplicated). The average size of the noncoding/nonrepeat portions of these blocks is 3.5 kb. All possible alignment combinations (293 pairs) of duplicated blocks containing V1R ORFs were analyzed for percent substitution. The resulting pairwise nucleotide divergence of the noncoding portions of the blocks is plotted versus the nucleotide divergence of the V1R coding sequences within the blocks. The right axis converts percent divergence into time, using a molecular clock rate of 0.5%/million years (Myr) for noncoding rodent sequences. The thicker black line between 20–29 Myr indicates the approximate date of mouse–rat speciation (36). The 252 possible comparisons of the NC/BD/BC subfamily (23 genes) are plotted as open blue circles, the 15 possible comparisons of the A-subfamily blocks (six genes) are plotted as red diamonds, the 21 possible comparisons of the B-subfamily blocks (seven genes) are plotted as green diamonds, the *V1ra7-V1ra8* comparison (two genes) is plotted as a green triangle, and the three *13.3–13.4–13.8* (three genes) comparisons and the *13.6–13.9* comparison (two genes) are plotted as gray circles.

The 5′ UTRs and regions immediately upstream of transcription start sites were examined for common regulatory features. Two patches of upstream homology among V1R paralogs at the 6D locus were revealed by PIPMAKER (Fig. 1*A*, and Fig. 7, which is published as supporting information on the PNAS web site). Except for these two conserved regions, the noncoding sequence around members of the different subfamilies at the 6D locus is unalignable. One patch surrounds the first noncoding exon of the four genes analyzed by RACE. A strong eukaryotic promoter is predicted here for three of the four genes. The sequence of this patch is highly conserved in the 5′ regions of all other V1R genes with ORFs at the 6D locus (Figs. 1*A* and 4), except V1R3b (see below). These regions exhibit comparable or greater homology as the coding regions, suggesting that these putative regulatory domains are under a high degree of selection. Two additional copies of this region are found not associated with V1R genes (Fig. 1*A*).

Among the ORFs at the 6D locus, the *V1Rb3* gene uniquely lacks the patch of promoter homology. *V1Rb3* probably arose by means of block duplication of *V1Rb1*, but the *V1Rb3* block is truncated ≈4 kb upstream of the coding sequence. Because the *V1Rb1* RACE product extends ≈2 kb beyond this point, *V1Rb3*, despite having an ORF, may be stranded without a promoter. Thus far, we have not been able to identify a *V1Rb3* cDNA.

Notably, these shared promoter motifs are specific to the 6D locus. The V1Rs at the other chromosome 6 locus may have their own common locus-specific promoter homology, however. Discrete regions upstream of V1Rs from this other locus show a higher degree of homology than the surrounding territory (Fig. 1*B*). For example, a region 5.0–5.5 kb upstream of NC2 is 80–90% homologous to other regions in the same cluster, whereas adjacent sequences are less similar (55–90%). Interestingly, we also identify two regions of noncoding homology shared by the two chromosome 6 V1R loci (Fig. 1).

Very few TRANSFAC database hits are universally conserved at common positions within either of the putative regulatory regions in the 6D locus. Two of the more noteworthy findings are a high incidence of Ikaros (IK2) motifs in the 5′ regions (9.2 hits/kb) and Lmo2 binding motifs in the 3′ regions (4.3 times more than in random control sequences with equivalent base pair composition). IK2 is a transcription factor expressed in myeloid progenitors (34), and Lmo2 is part of a transactivating complex believed to play a role in hematopoiesis (35). Another intriguing connection to hematopoietic regulation is a ≈150-bp region downstream of *V1Ra7* (and at the 3′ end of each of the other three *V1Ra7*-like duplicated blocks). This region is 87% identical to the promoter region of the mouse Fc gamma IIIA gene and includes a ≈40-bp region that is 90% identical to sequences near the 5′ DNaseI-hypersensitive site in the mouse β-globin locus control region.

Discussion

The V1R receptors expressed in the sensory neurons of the murine VNO are involved in the detection of pheromones (1, 2). We conducted a comparative genomics study of the mouse V1R gene family to address two general questions. First, because pheromonal systems contribute to species-specific behaviors, how are V1R repertoires changing in evolution to adapt to species-specific requirements? Second, because the ability to respond appropriately to pheromones is based on an organizing principle in which individual sensory neurons transcribe only a single pheromone receptor, what is the molecular mechanism that assures exclusive expression of V1R genes? Our analyses of

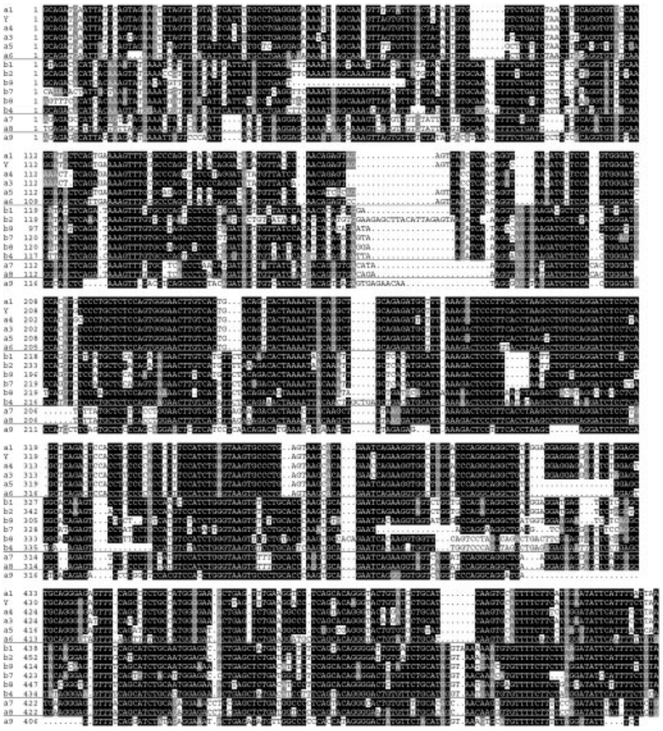

Fig. 4. Common V1R promoter homology at the chromosome 6D locus. The common promoter regions of the 15 V1R genes with ORFs from the chromosome 6D locus are aligned (these regions correspond to the purple-shaded regions in Fig. 1*A*). The 15 V1R genes belong to four divergent subfamilies separated by lines. Identities are shaded black; nucleotide positions shared by at least three V1Rs are shaded gray. See Fig. 9, which is published as supporting information on the PNAS web site, for an enlarged version of this figure.

three V1R loci in the mouse genome reveals a history of recent and frequent gene duplication events that have led to significant changes in the mouse V1R repertoire over a short period of evolution. In addition, we identify a common putative regulatory structure shared by V1Rs within one cluster.

V1R Expansion: A Role in Reinforcing Speciation? Species-specific responses to pheromones are likely to be reflected in genomic differences among VNO receptor genes. Our analyses indicate that gene-block duplications occurred in recent rodent evolution to significantly expand the murine V1R repertoire. In the three loci examined, ≈7 ancestral prototypes gave rise to 73 genes, of which 44 are potentially functional. These V1R duplications can be dated approximately by using a molecular clock rate for nucleotide substitutions in noncoding rodent sequences (25). V1R expansion at all three loci correlates with a major speciation event (Fig. 3). According to the average divergences between duplicated blocks, these V1Rs multiplied predominantly ≈20–30 million years ago. This period overlaps the period when mouse and rat diverged, as assessed by similar molecular clock methods (36).

Lx, but not L1_MM, L1 elements were included in some of the V1R block duplications. Thus, these duplications occurred between the periods of Lx and L1_MM activity in the mouse genome. It has been established that Lx repeats were active during and just before the mouse–rat split, and L1_MM repeats were active just after rat diverged from mouse (24). Therefore, the V1R duplications can be dated very close to the mouse–rat split by the independent measure of murine L1 insertions.

The transition from non-V1R to V1R territory at the 6D locus is marked by a striking increase in repeat density (Fig. 1*A*). The prevalence of L1 and long dinucleotide repeats between duplicated blocks raises the possibility that these repeats played a role in the rearrangements leading to V1R expansion. Simple repeat structures may be favored targets of transposition (37) and could have favored the numerous L1 insertions into these regions. Frequent L1 insertions imply a history of frequent double-strand breaks and an increased propensity for rearrangement. Moreover, L1 sequences, once inserted on either side of a gene block, could mediate unequal homologous recombination events, leading to changes in gene number.

It is noteworthy that most of the L1 activity in these regions coincides roughly with the mouse–rat split and the duplication events. Approximately half of the L1 elements between the duplicated blocks belong to the L1_MM or Lx subfamilies. Moreover, L1_MMs are especially abundant at the 6D V1R locus where most duplications date just after or during mouse–rat divergence, whereas Lx elements are especially abundant at the other two V1R loci where most duplications date just before or during mouse–rat divergence (Figs. 1 and 3). Although this association could be coincidental, it is tempting to speculate that the activities of these repeat subfamilies played a role in the V1R gene duplications.

Taken together, our data favor a model in which changes in V1R gene content, perhaps caused by surrounding repeats, led directly to marked changes in pheromone recognition around the time mouse and rat diverged. These changes in the V1R repertoire may have contributed to the reproductive barriers that separated mouse and rat during speciation. A similar study of orthologous V1R loci in rat and other rodents will be informative. We would expect to find duplications, and perhaps local repeat activity, that resulted in species-specific changes in functional V1R content.

The human V1R subgenome is very different from that of mice. We find no extensive clustering of human V1R genes. We find no V1R homologs of the mouse chromosome 6D locus at the expected syntenic location in the human genome. Almost all human V1R homologs are pseudogenes (ref. 27 and results herein). Although there is anecdotal evidence for human pheromonal function (38), no molecular or physiological basis has yet been described for this form of human communication. Moreover, the Trp2 ion channel, thought to play a major role in the transduction of pheromone binding in the mouse VNO (ref. 39; B. Leypold, R. Yu, and R.A., unpublished work), is a pseudogene in humans (40). It also has been difficult to identify a distinct fiber tract projecting from the VNO to the brain in human specimens (41). Thus, a loss of selective pressure may have contributed to the loss of function of the human pheromone response systems.

Alternatively, if speciation is accompanied by significant changes in pheromone receptor repertoires, human receptors might not resemble those used by rodents. One potentially functional human V1R-like coding sequence has been identified (ref. 28 and results herein), and it is less than 30% identical at the amino acid level to the closest known mouse V1R. It maps to human chromosome 19q13.4, but the syntenic region on mouse chromosome 7 contains no known V1R-like genes. Thus, if functional human pheromonal receptors exist, they share minimal orthology to rodent V1R repertoires. It remains to be seen whether the decline in V1R homology in the human genome is caused by loss of pheromonal function or functional differences that warrant unique receptor repertoires in mice and human.

A Locus-Specific Regulatory Structure. Mouse V1R genes exhibit many common aspects of transcriptional regulation—they are expressed in the same neuronal cell type and zone of the VNO and at the same time during development (8, 9). They also share transcriptional mechanisms that assure allelic exclusion (14). We have identified putative promoter regions for V1R genes at the mouse 6D locus. Although they represent four divergent subfamilies, the 15 potentially functional V1R genes at this locus have two conserved $\approx$1-kb noncoding structures (Fig. 1*A*). One of these regions is typically found within the 5′ UTR. This region could be important for regulation at the level of translational efficiency, subcellular targeting, or transcript stability. The second region of homology contains strong RNA polymerase II promoter motifs and is typically found $\approx$5 kb upstream of the coding regions (Fig. 1*A*). Transcription start sites identified empirically for four V1R genes map to this region of homology. This homology likely represents conservation of regulatory motifs that may bind transcription factors that specify the appropriate time and place for expression.

The remarkable conservation of these upstream promoter regions among otherwise divergent V1R genes (Fig. 4) is unusual both in its length and degree of conservation. Promoter regions typically are scattered with short, degenerate transcription factor-binding sites (42). The highly conserved sequences upstream of V1R genes may compete for a limiting transcriptional assembly or represent a conserved structure for V1R gene coregulation, perhaps mediated by an locus control region. Interestingly, we identified regions of homology to the β-globin locus control region DNase hypersensitive sites, and these sequences are candidate control regions. We also find isolated regions of V1R promoter homology not apparently associated with V1R genes or local duplication (Fig. 1*A*). These orphan promoter regions are candidate enhancers. Finally, because these conserved structures are found only within the 6D V1R cluster, additional regulatory features must exist that could be involved with a higher level of locus choice. Specific regions of homology detected between the two chromosome 6 loci (Fig. 1) are candidates for cross-locus regulation. Further experiments aimed at elucidating the significance of these apparent locus-specific regulatory structures and cross-locus homologies should shed light on how V1Rs are coregulated such that one, and only one, gene is expressed per cell.

We thank A. Smit for help analyzing repeat structures; C. Friedman for mouse fluorescence *in situ* hybridization experiments; T. Newman for help with graphical depictions of repeats; E. Williams and G. van den Engh for help in assessing the burst-like characteristics of V1R expansion; B. Leypold for providing a cDNA library for RACE amplification from mouse VNO tissue; E. Davidson, J. Young, and L. Belluscio for feedback and stimulating dialogue; and members of the core sequencing facility in the University of Washington Department of Molecular Biotechnology for technical support. A portion of these data were generated through use of the Celera Discovery System and Celera Genomics' associated databases. This work was supported by the Howard Hughes Medical Institute and National Institutes of Health Grant R01-DC04209.

1. Keverne, E. B. (1999) *Science* **286,** 716–720.
2. Krieger, J., Schmitt, A., Lobel, D., Gudermann, T., Schultz, G., Breer, H. & Boekhoff, I. (1999) *J. Biol. Chem.* **274,** 4655–4662.
3. Aujard, F. (1997) *Physiol. Behav.* **62,** 1003–1008.
4. Curtis, J. T., Liu, Y. & Wang, Z. (2001) *Brain Res.* **901,** 167–174.
5. Buck, L. & Axel, R. (1991) *Cell* **65,** 165–187.
6. Sullivan, S. L., Adamson, M. C., Ressler, K. J., Kozak, C. A. & Buck, L. B. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 884–888.
7. Xie, S. Y., Fienstein, P. & Mombaerts, P. (2000) *Mamm. Genome* **11,** 1070–1078.
8. Dulac, C. & Axel, R. (1995) *Cell* **83,** 195–206.
9. Pantages, E. & Dulac, C. (2000) *Neuron* **28,** 835–845.
10. Herrada, G. & Dulac, C. (1997) *Cell* **90,** 763–773.
11. Ryba, N. J. & Tirindelli, R. (1997) *Neuron* **19,** 371–379.
12. Matsunami, H. & Buck, L. B. (1997) *Cell* **90,** 775–784.
13. Chess, A., Simon, I., Cedar, H. & Axel, R. (1994) *Cell* **78,** 823–834.
14. Rodriguez, I., Feinstein, P. & Mombaerts, P. (1999) *Cell* **97,** 199–208.
15. Trask, B. J. (1999) in *Genome Analysis: A Laboratory Manual*, eds. Birren, B., Green, E., Heiter, P., Klapholz, S., Myers, R., Reithman, H. & Roskams, J. (Cold Spring Harbor Lab. Press, Plainview, NY), Vol. 4, pp. 303–413.
16. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. (1998) *Genome Res.* **8,** 175–185.
17. Ewing, B. & Green, P. (1998) *Genome Res.* **8,** 186–194.
18. Gordon, D., Abajian, C. & Green, P. (1998) *Genome Res.* **8,** 195–202.
19. Del Punta, K., Rothman, A., Rodriguez, I. & Mombaerts, P. (2000) *Genome Res.* **10,** 1958–1967.
20. Schwartz, S., Zhang, Z., Frazer, K. A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R. & Miller, W. (2000) *Genome Res.* **10,** 577–586.
21. Quandt, K., Frech, K., Karas, H., Wingender, E. & Werner, T. (1995) *Nucleic Acids Res.* **23,** 4878–4884.
22. Belluscio, L., Koentges, G., Axel, R. & Dulac, C. (1999) *Cell* **97,** 209–220.
23. Smit, A. (1999) *Curr. Opin. Genet. Dev.* **9,** 657–663.
24. Furano, A. V., Hayward, B. E., Chevret, P., Catzeflis, F. & Usdin, K. (1994) *J. Mol. Evol.* **38,** 18–27.
25. Li, W.-H., Luo, C. & Wu, C. (1985) in *Molecular Evolutionary Genetics*, ed. Macintyre, R. J. (Plenum, New York), pp. 1–94.
26. Ngai, J., Dowling, M. M., Buck, L., Axel, R. & Chess, A. (1993) *Cell* **72,** 657–666.
27. Giorgi, D., Friedman, C., Trask, B. J. & Rouquier, S. (2000) *Genome Res.* **10,** 1979–1985.
28. Rodriguez, I., Greer, C. A., Mok, M. Y. & Mombaerts, P. (2000) *Nat. Genet.* **26,** 18–19.
29. Sosinsky, A., Glusman, G. & Lancet, D. (2000) *Genomics* **70,** 49–61.
30. Tsuboi, A., Yoshihara, S. Yamazaki, N., Kasai, H., Asai-Tsuboi, H., Komatsu, M., Serizawa, S. Ishii, T., Matsuda, Y., Nagawa, F. & Sakano, H. (1999) *J. Neurosci.* **19,** 8409–8418.
31. Bulger, M., Bender, M. A., van Doorninck, J. H., Wertman, B., Farrell, C. M., Felsenfeld, G., Groudine, M. & Hardison, R. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 14560–14545.
32. Lane, R. P., Cutforth, T., Young, J., Athanasiou, M., Friedman, C., Rowen, L., Evans, G., Axel, R., Hood, L. & Trask, B. J. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 7390–7395.
33. Lane, R. P., Roach, J., Lee, I. Y., Boysen, C., Smit, A., Trask, B. J. & Hood, L. (2001) *Genome Res.*, in press.
34. Koipally, J., Renold, A., Kim, J. & Georgopoulos, K. (1999) *EMBO J.* **18,** 3090–3100.
35. Wadman, I. A., Osada, H., Grutz, G. G., Agulnick, A. D., Westphal, H., Forster, A. & Rabbitts, T. H. (1997) *EMBO J.* **16,** 3145–3131.
36. O'hUigin, C. & Li, W. H. (1992) *J Mol. Evol.* **35,** 377–384.
37. Tatout, C., Lavie, L. & Deragon, J. M. (1998) *J. Mol. Evol.* **47,** 463–470.
38. Stern, K. & McClinktock, M. K. (1998) *Nature (London)* **392,** 177–179.
39. Liman, E. R., Corey, D. P. & Dulac, C. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 5791–5796.
40. Vannier, B., Peyton, M., Boulay, G., Brown, D., Qin, N., Jiang, M., Zhu, X. & Birnbaumer, L. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 2060–2064.
41. Trotier, D., Eloit, C., Wassef, M., Talmain, G., Bensimon, J. L., Doving, K. B. & Ferrand, J. (2000) *Chem. Senses* **25,** 369–380.
42. Kirchhamer, C. V., Yuh, C. H. & Davidson, E. H. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 9322–9328.