

Research article

Open Access

The G protein-coupled receptor subset of the rat genome

David E Gloriam*, Robert Fredriksson and Helgi B Schiöth*

Address: Department of Neuroscience, Uppsala University, BMC, Box 593, 751 24, Uppsala, Sweden

Email: David E Gloriam* - DavidGloriam@googlemail.com; Robert Fredriksson - robert.fredriksson@neuro.uu.se;

Helgi B Schiöth* - helgis@bmc.uu.se

* Corresponding authors

Published: 25 September 2007

Received: 13 April 2007

BMC Genomics 2007, 8:338 doi:10.1186/1471-2164-8-338

Accepted: 25 September 2007

This article is available from: <http://www.biomedcentral.com/1471-2164/8/338>

© 2007 Gloriam et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The superfamily of G protein-coupled receptors (GPCRs) is one of the largest within most mammals. GPCRs are important targets for pharmaceuticals and the rat is one of the most widely used model organisms in biological research. Accurate comparisons of protein families in rat, mice and human are thus important for interpretation of many physiological and pharmacological studies. However, current automated protein predictions and annotations are limited and error prone.

Results: We searched the rat genome for GPCRs and obtained 1867 full-length genes and 739 pseudogenes. We identified 1277 new full-length rat GPCRs, whereof 1235 belong to the large group of olfactory receptors. Moreover, we updated the datasets of GPCRs from the human and mouse genomes with 1 and 43 new genes, respectively. The total numbers of full-length genes (and pseudogenes) identified were 799 (583) for human and 1783 (702) for mouse. The rat, human and mouse GPCRs were classified into 7 families named the *Glutamate*, *Rhodopsin*, *Adhesion*, *Frizzled*, *Secretin*, *Taste2* and *Vomerinasal1* families. We performed comprehensive phylogenetic analyses of these families and provide detailed information about orthologues and species-specific receptors. We found that 65 human *Rhodopsin* family GPCRs are orphans and 56 of these have an orthologue in rat.

Conclusion: Interestingly, we found that the proportion of one-to-one GPCR orthologues was only 58% between rats and humans and only 70% between the rat and mouse, which is much lower than stated for the entire set of all genes. This is in mainly related to the sensory GPCRs. The average protein sequence identities of the GPCR orthologue pairs is also lower than for the whole genomes. We found these to be 80% for the rat and human pairs and 90% for the rat and mouse pairs. However, the proportions of orthologous and species-specific genes vary significantly between the different GPCR families. The largest diversification is seen for GPCRs that respond to exogenous stimuli indicating that the variation in their repertoires reflects to a large extent the adaptation of the species to their environment. This report provides the first overall roadmap of the GPCR repertoire in rat and detailed comparisons with the mouse and human repertoires.

Background

The rat genome [1] was the third mammalian genome to be sequenced, after the human [2,3] and mouse [4] genomes. The rat is one of the most widely used model organisms in biological research. Despite this fact, the annotated proteins are fewer for rat (26,123) compared to mouse (30,397) and human (41,973) (Ensembl "peptides known" and "peptides known-ccds" datasets March 2007) [5]. Automated protein predictions offer fast annotation but they are error-prone and need to be followed up by careful manual curation. For instance the Genscan gene prediction program has a sensitivity and specificity of about 90% for detecting exons, leading to frequent errors in multi-exon genes [6]. Our recent annotation of the chicken G protein-coupled receptors (GPCRs) showed that over 60% of the chicken Genscan gene predictions with a human orthologue needed curation. The curation drastically increased the quality of the dataset as the average percentage identity between the human-chicken one-to-one orthologous pairs was raised from 56% to 73% [7]. Moreover, automated annotation based on inter-species comparisons has difficulties in identifying species-specific proteins, a problem that is relevant especially to large protein families that display large differences between species. For example previous reports of the V1R and V2R vomeronasal receptor repertoires display a very large variation in numbers [8-11]. The degrees of completeness and accuracy of protein annotations have, of course, a substantial impact on subsequent analyses such as phylogenetic analyses and calculations of evolutionary distances. Accurate comparisons of the rat and human proteins, such as correct assignment of orthologous pairs, are crucial for the design and interpretation of physiological and pharmacological studies in which results are inferred between the species.

The superfamily of GPCRs is one of the largest groups of proteins within most mammals. GPCRs are signal mediators that have a prominent role in most major physiological processes at both the central and peripheral level [12]. It has been estimated that about 80% of all known hormones and neurotransmitters activate cellular signal transduction mechanisms via GPCRs [13]. The key common structural components of the GPCRs are the seven transmembrane α -helices that span the cell membrane. It has been estimated that GPCRs represent between 30-45% of the current drug targets [14,15]. However, drugs have only been developed for a very small number of the GPCRs and the potential for further drug discovery within this field is enormous.

The human GPCR repertoire has previously been divided into five main families (GRAFS); *Glutamate* (clan C), *Rhodopsin* (clan A, includes the olfactory receptors), *Adhesion* (clan B2), *Frizzled/Taste2* and *Secretin* (clan B) [16]. The

GRAFS families are found in all bilateral species and it has thus been suggested that they arose before the split of nematodes from the chordate lineage [17]. The largest family is by far the *Rhodopsin* family, one reason being that it includes the many olfactory receptors (ORs). Most of the GPCR drug targets, mainly amine and peptide receptors, are found within this family [18]. In humans, the second largest family is the *Adhesion* family. Most of the receptors in this family are still orphans (without a known ligand) [19,20]. The *Glutamate* family includes receptors that bind to glutamate, GABA and calcium as well as the groups of sweet and umami taste receptors (TAS1Rs) and vomeronasal receptors type 2 (V2Rs) that are binding native exogenous compounds. The *Secretins* bind large peptides such as secretin, parathyroid hormone, glucagon, glucagon-like peptide, calcitonin, vasoactive intestinal peptide, growth hormone releasing hormone and pituitary adenylyl cyclase activating protein. The *Frizzled* receptors bind among others the Wnt ligand and play an important role in embryonic development. The vomeronasal receptors type 1 (*Vomeronal1*) and taste receptor type 2 (*Taste2*) GPCR families are involved in pheromone recognition and bitter taste sensing, respectively. Moreover, there are gene sequences that have been suggested to represent additional GPCRs but show only weak or no sequence similarity to any of the main GPCR families.

Here we provide the first overall map of the GPCR subset of the rat genome. We have made extensive efforts to mine the GPCR repertoire and cover all GPCR families comprising a total of 1867 (+739 pseudogenes) gene sequences. In addition we have also updated the human and the mouse repertoires. We have performed phylogenetic analyses of the rat, human and mouse GPCR families and provide detailed information of their orthologous relationships.

Results

The overall GPCR repertoires in rat, mouse and human

In this analysis we present the overall repertoire of GPCRs in rats and updated versions of the GPCR repertoires in humans and mice. The GPCR families found in humans, GRAFS, have been slightly updated since they were last described by our group, but the mouse datasets are nearly identical [21]. The main difference is that in this analysis the *Vomeronal1* family and the olfactory receptor (OR) group are included and the *Frizzled* and *Taste2* receptors are considered as two separate families. Throughout this article the GPCR families are written in italics; *Adhesion*, *Frizzled*, *Glutamate*, *Rhodopsin*, *Secretin*, *Taste2* and *Vomeronal1*, whereas the names of family subgroups are abbreviated e.g. V2R (vomeronasal type 2 receptors) and OR (olfactory receptors). Our classification of the GPCR families is based strictly on sequence homology. This analysis also includes 21 gene sequences that have been

suggested to encode GPCRs but do not belong to any of the GPCR families (i.e. they have no sequence similarity). These are referred to as "Other GPCRs" or just "Others". The numbers of GPCR genes in the different families and species are given in Table 1 and the names of the receptors in each family and species is listed in a separate spreadsheet [see Additional File 1]. In Table 1, the GPCR genes have been categorised as "full-length", "pseudogenes" or "partial" and the numbers of new genes are also indicated. Full-length GPCR genes contain an intact transmembrane domain and are likely to encode functional receptor proteins. The pseudogenes are non-functional genes that have been frame-shifted due to additions or deletions of nucleotides or truncated by missense mutations introducing stop codons. The "partial" genes are missing parts of their sequence due to incomplete information in the respective genome assembly and in sequence databases. GPCR gene sequences were defined as new if they had not been annotated as GPCRs in Genbank and did not exist in any of the supplementary datasets linked to previous studies. The GPCR repertoires in the rodents are more than twice as large as that in humans, and rats have more GPCRs than mice. This is mainly explained by the large differences in the number of ORs, but a substantial part of the human-rodent differences is because of the vomeronasal receptors, *Vomeronasal1s* and V2Rs (a subset within the Glutamate family). There are no partial human GPCR genes, 2 in mice and 38 (1.4% of all genes) in rats. These figures reflect the completeness of the respective genome assemblies which must now all be considered fairly complete. Most (28/38) of the partial rat sequences are OR genes. These are difficult to assemble due to their large numbers, high sequence similarity and genomic proximity. In this respect this analysis shows an important difference as compared with for example the initial work on mouse by Vassilatis and colleagues [22], which contained large numbers of partial sequences. Our sequence datasets [see Additional files 2, 3, 4] have, in all instances possible, been carefully compared to previously published datasets (see below).

Phylogenetic analyses

We produced phylogenetic trees for each GPCR family. These are shown in Figures 1, 2, 3 and as additional files of this article [see Additional files 5, 6, 7], which all also hold pie charts of the proportions of orthologous and species-specific GPCRs in the three species analysed. The mouse *Glutamate*, *Rhodopsin* (non-ORs), *Adhesion*, *Frizzed* and *Secretin* GPCRs are not included in these trees since they have been described before in relation to the human repertoires [21], and the exact orthologous and paralogous relationships are viewed in a table listing the receptors in the three species studied [see Additional File 1]. All receptors sequences were trimmed prior to the phylogenetic analysis to isolate the transmembrane domains

which are the regions that are conserved throughout the GPCR families. A special approach was used for the phylogenetic analysis of the *Rhodopsin* family which is very challenging because of its size and diversity. Some *Rhodopsin* GPCRs have diverged to the extent that their relationships to the rest of the family members cannot be determined. These receptor sequences do not show stable tree topology and group with different subfamilies when using different phylogenetic algorithms. The ambiguous grouping of these receptors is also illustrated by the fact that when their sequences are searched against the whole dataset using BLAST the best matches (e.g. top 5) to these receptors belong to different subfamilies. We minimised the above problems by separating the receptors with a low sequence identity from the main part of the dataset when performing phylogenetic analyses. This was done by applying iteratively higher thresholds of percentage sequence identity on the dataset until a stable tree topology could be obtained.

The Rhodopsin family

The phylogenetic tree(s) of the *Rhodopsin* family in human and rat is shown in Figure 1 and 2. The majority of the *Rhodopsins* are in Figure 1, whereas other members with ambiguous relationships are shown in Figure 2 as separate groups, pairs and singles in the order of falling sequence identity (to the sequences before them). Sequences defined by the same percentage of sequence identity are grouped in boxes with the threshold value given in the lower right corner. In Figure 1 the α - δ groups, previously defined by our group [16] were reproduced and they contain most of the receptors from the original analysis. 79% of the human and rat *Rhodopsins* are found in one-to-one orthologous pairs.

Rhodopsin family species variation

Several of the *Rhodopsin* family GPCRs show interesting species variation. Most of the species-specific *Rhodopsins* are found among the MAS-related GPCRs (MRGs), trace-amine-associated receptors (TAARs) and "formyl peptide receptor-like" (FPRLs) receptors. The group of MRGs are located primarily in sensory neurons of the dorsal root ganglion and are implicated in perception of pain [23]. The numbers of MRGs (at the 34% threshold) are 10, 15 and 28 in human, rat and mouse, respectively. There are 5 rat-human orthologous pairs of MRG receptors. The TAARs (in the cluster of amine binding receptors) are 5, 15 and 17 in human, mouse and rat, respectively [24,25]. The FPRLs (slightly below and to the left of the centre of the largest tree) *Fprrs2*, *Fprrs3*, *Fprrs4* and *Fprrs6* are rat specific, whereas *FPRL2* is only found in humans. The remaining species-specific *Rhodopsins* are singles (shown in bold style in Figure 1, 2) derived from a single gene duplication or gene loss in either the primate or rodent lineage. These are: 5-HT1E and 5-HT5B serotonin recep-

Table 1: The human, rat and mouse GPCR family repertoires

GPCR Family	Human					Rat						Mouse					
	Full-length	Pseudo-genes	Pseudo-genes (%)	New full-length	New pseudo-genes	Full-length	Pseudo-genes	Partial	Pseudo-genes (%)	New full-length	New pseudo-genes	Full-length	Pseudo-genes	Partial	Pseudo-genes (%)	New ¹ full-length	New ¹ pseudo-genes
<i>Adhesion</i>	33	2	6%	0	2	26	1	4	3%	0	2	30	1	3%	0	0	
<i>Frizzled</i>	11	0	0%	0	0	10	1		9%	0	0	11	0	0%	0	0	
<i>Glutamate</i>	22	0	0%	0	0	22	0		0%	0	0	22	0	0%	0	0	
<i>-V2R subgroup</i>	1	11	92%	1	11	108	87	2	45%	23	53	111	183	1	62%	21	63
<i>Rhodopsin</i>	284	27	9%	0	1	297	7	4	2%	5	0	320	23	7%	3	NA*	
<i>-OR subgroup</i>	388	479	55%	0	12	1234	552	28	31%	1234	552	1081	325	1	23%	NA*	NA*
<i>Secretin</i>	15	0	0%	0	0	15	0		0%	0	0	15	0	0%	0	0	
<i>Taste2</i>	25	10	29%	0	0	35	6		15%	0	0	35	5	13%	0	0	
<i>Vomeronal I</i>	5	53	91%	0	53	105	84		44%	12	76	145	164	53%	17	92	
<i>Other GPCRs¹</i>	15	1	6%	0	1	15	1		14%	1	0	13	1	7%	1	0	
Total	799	583		1	133	1867	739	38		1276	759	1783	702	2	43	247	

The human, rat and mouse GPCR families; their numbers of full-length, pseudogene and partial members; as well as the number of new members identified in this study. GPCR gene sequences were defined as new if their sequence was not available in a public database annotated as a GPCR. *No figure could not be calculated.

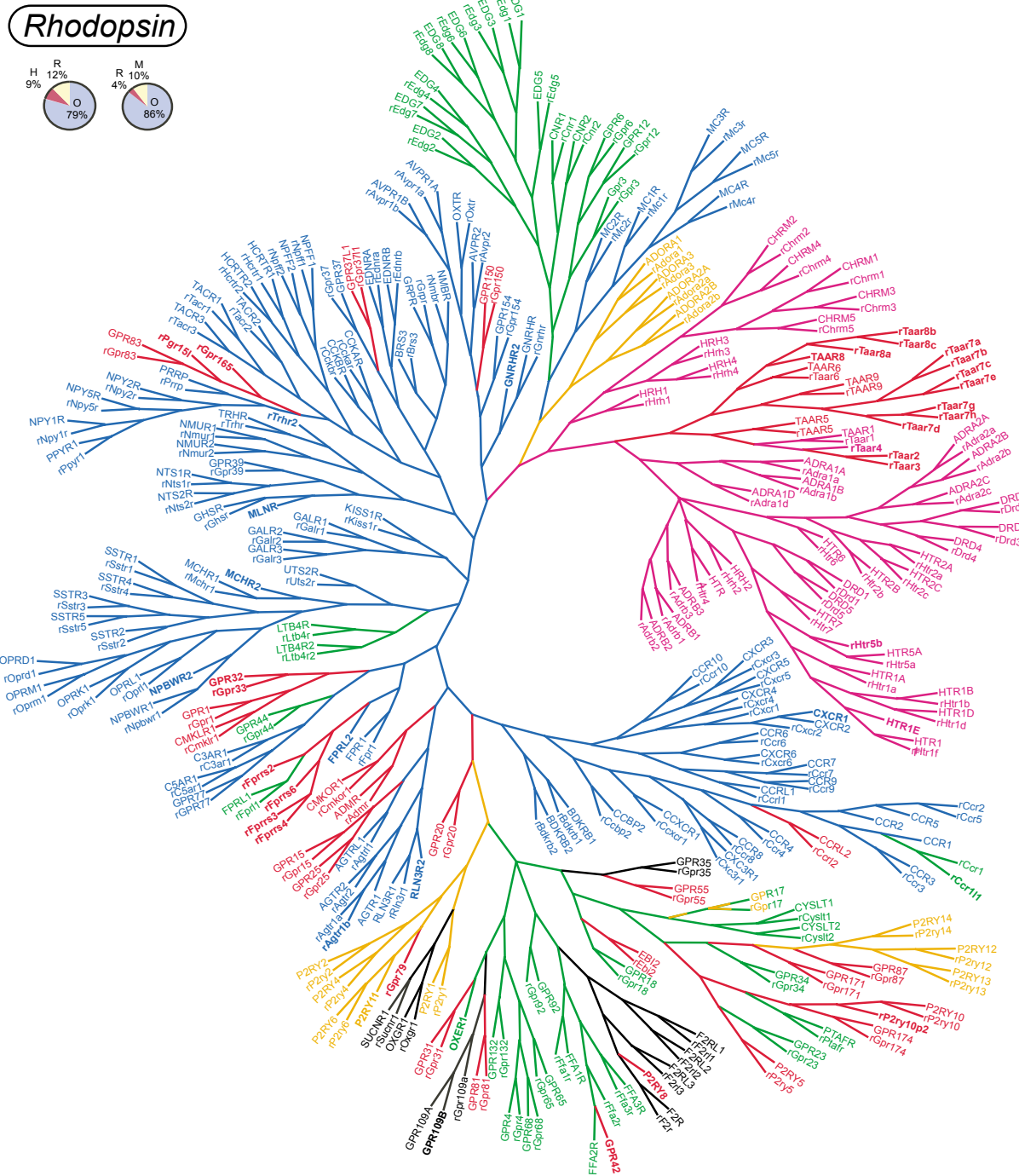


Figure 1
Phylogenetic analysis of the Rhodopsin family. The figure shows a phylogenetic tree of Rhodopsin family receptors. Rhodopsins which display ambiguous relationships were analysed in separate (see Figure 2). Species-specific receptors have names in bold style. The ligand types of the receptors are indicated with the following colours; red: orphan, blue: peptide, lilac: amine, green: lipid-like, brown: purine, turquoise: opsin and black: other. The first pie chart above the trees shows the proportions of human-rat one-to-one orthologues (O), human specific (H) and rat specific (R) members. The second pie chart displays the proportions of rat-mouse one-to-one orthologues (O), rat specific (R) and mouse specific (M) members. This phylogenetic tree is a consensus tree of 2 consensus trees derived from 100 maximum parsimony and neighbour joining analyses, respectively, and calculated using the UNIX version of the Phylip 3.6 package [73].

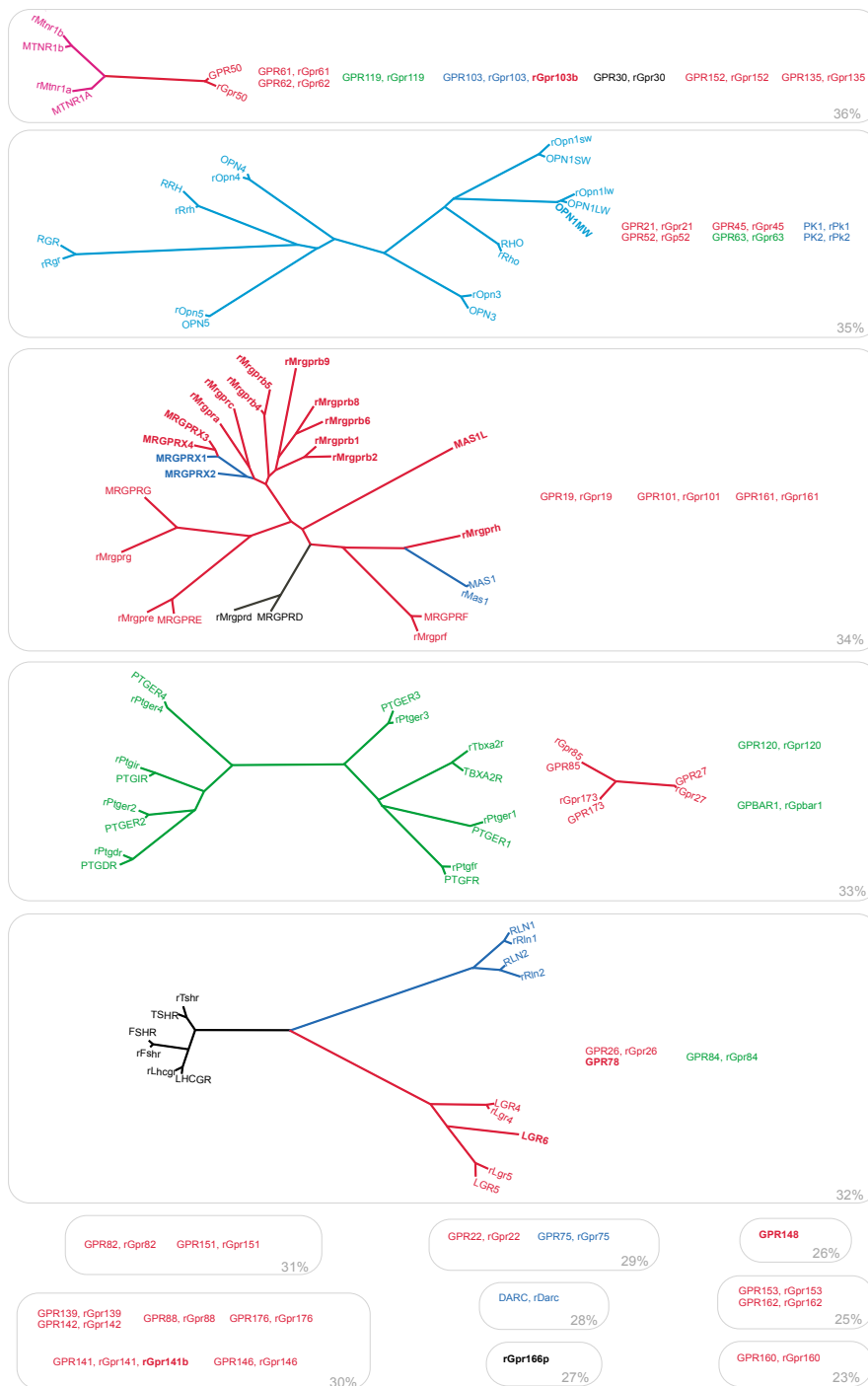


Figure 2
Rhodopsin family receptors with ambiguous relationships to other members. Groups, pairs and "single" Rhodopsin family receptors that display ambiguous relationships to other members (see material and methods). These are shown in order of decreasing sequence identity (given in the lower right corner of each box). Species-specific receptors have names in bold style. The ligand types of the receptors are indicated with the following colours; red: orphan, blue: peptide, lilac: amine, green: lipid-like, brown: purine, turquoise: opsin and black: other. The phylogenetic trees are consensus trees of 100 maximum likelihood trees for which branch lengths were calculated using TreePuzzle (see material and method). The olfactory subfamily/ group of the Rhodopsin family was analysed in a separate phylogenetic analysis [see Additional files 7 and 8].

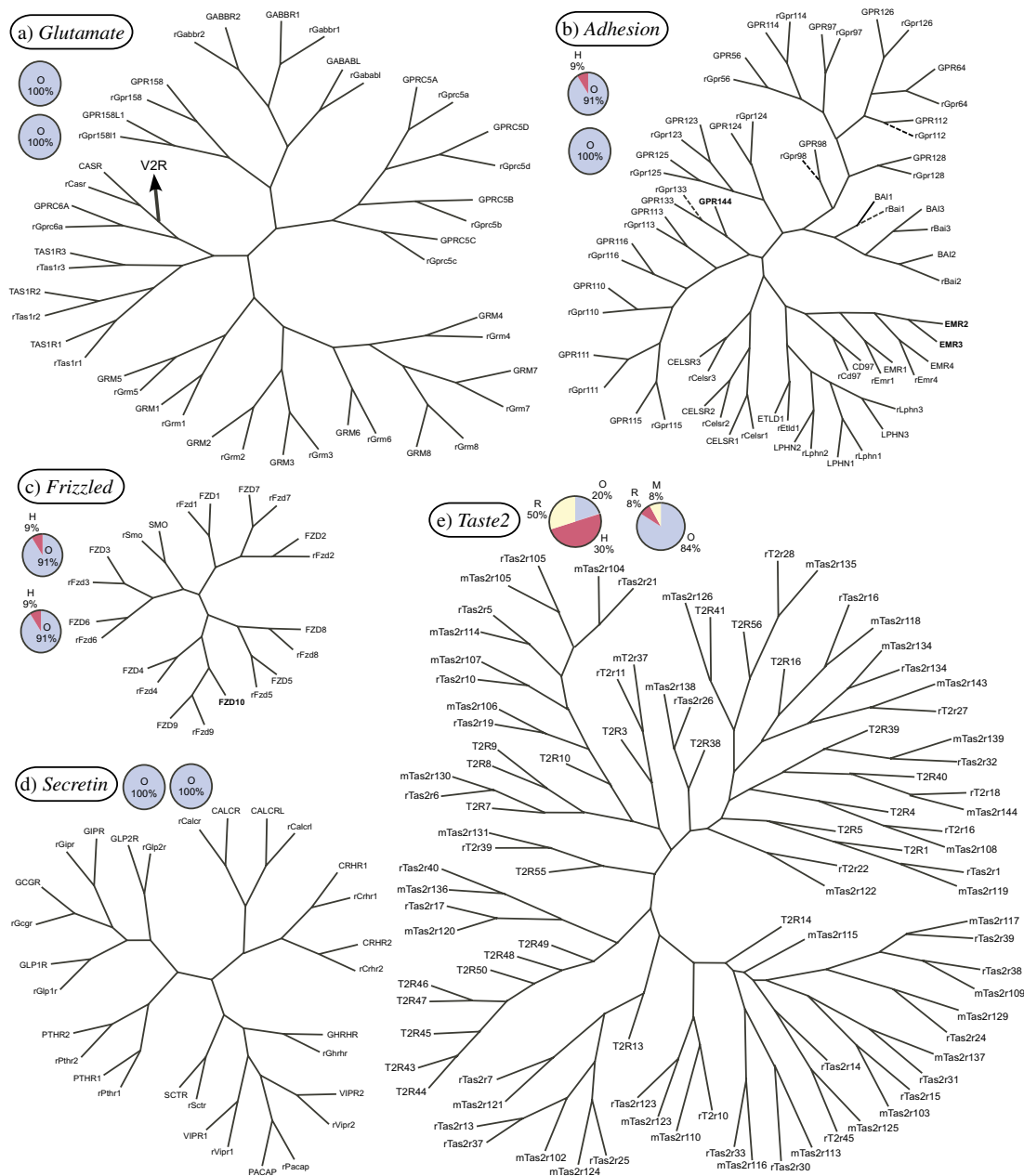


Figure 3
Phylogenetic trees of the Glutamate, Adhesion, Frizzled, Secretin and Taste2 families. The figure shows the consensus trees of 100 maximum parsimony trees of the human and rat a) *Glutamate*, b) *Adhesion*, c) *Frizzled*, d) *Secretin* and e) *Taste2* (mouse included) GPCR families. The first pie chart after the each family name shows the proportions of human-rat one-to-one orthologues (O), human specific (H) and rat specific (R) members. The second pie chart displays the proportions of rat-mouse one-to-one orthologues (O), rat specific (R) and mouse specific (M) members. In Figure 3, the place of the large group of V2Rs is indicated with an arrow. Rat sequences that were excluded from the phylogenetic analysis because they were incomplete (due to gaps in the genome assembly) are displayed with dashed branches.

tors (HTR1E and Htr5b; among the amine binding receptors); Gonadotropin releasing hormone receptor 2 (GnRHR2), Gpr165, Pgr15l, Thyrotropin releasing hormone receptor 2 (Trhr2) and the motilin receptor (MLNR) (among the many peptide receptors); MCHR2 and NPBWR2 (at the roots of the somatostatin and opioid receptors, respectively); GPR32 and Gpr33 (under the somatostatin and opioid receptor cluster in Figure 2); Agtr1b and RLN3R2 (left of the chemokine and purine receptor clusters); P2RY11, Gpr79, OXER1, GPR109B, GPR42, P2RY8, P2ry10p2 (among the purine and F2R receptor cluster); CXCR1, and Ccr11l (among the chemokine receptors); Gpr103b (36% threshold); OPN1MW (35% threshold); GPR78, LGR6 (32% threshold); Gpr141b (30% threshold); Gpr166p (27%) and GPR148 (26% threshold).

Rhodopsin family ligand types and orphan receptors

The colour coding of Figure 1 and 2 gives an overview of the types of ligands or bound proteins for the *Rhodopsin* family receptors. There are 100 peptide (blue), 45 lipid (green), 38 amine (lilac), 13 purine (brown) -binding human *Rhodopsins* and also 9 opsin receptors that are by light (turquoise). One receptor, GPR17 has a dual ligand specificity for purine and lipid molecules [26]. The remaining 14 *Rhodopsins* (black) are scattered in smaller groups of ligand types including protease-activated (F2R, F2RL1, F2RL2 and mF2RL3), glycoprotein hormone (FSHR, LHCGR and TSHR), metabolite (OXGR1, GPR35, SUCNR1, GPR109A and GPR109B), amino acid (MRG-PRD) and steroid hormone (estrogen receptor GPR30). Moreover, we found that there are 65 human orphan *Rhodopsins*, of which 56 have orthologues in rats. 27 of the human orphan *Rhodopsins* are grouped in the phylogenetic tree of the majority of the receptors (Figure 1), whereas 38 have more atypical sequence and are found in smaller groups, pairs or are singles (Figure 2). 27 of the human orphan *Rhodopsins* could not be grouped with any characterized receptor making ligand prediction difficult. Additional information about the phylogenetic grouping of ligand types and orphan receptors was presented in a recent study by Surgand et al [27].

The Glutamate, Adhesion, Frizzled, Secretin and Taste2 families

The *Glutamate* (except for the V2R group) (Fig. 2a) and *Secretin* (Fig. 2d) families have a completely conserved number of members between rats, mice and human and they are all one-to-one orthologues. GPCRs belonging to the Glutamate family have been shown to form functional heterodimers or homodimers. The GABAB receptor is formed by heterodimerization between one GABAB1 subunit and one GABAB2 subunit, the sweet taste receptor is formed by heterodimerization between the T1R2 subunit and the T1R3 subunit, whereas the umami taste receptor

is formed by heterodimerization between the T1R1 subunit and the T1R3 subunit. Additional GPCRs within the Secretin family are formed by heterodimerization between a GPCR and an accessory protein [28]. In the *Frizzled* family (Fig. 2c) one rat member, Fzd10, is a pseudogene. Two *Adhesions* (Fig. 2b), EMR2 and EMR3, were previously shown to be missing in mice [29] and these are also missing in the rat genome. Moreover, we now found that the rat and mouse gene sequences of Gpr144 contain stop codons within the transmembrane region and are thus pseudogenes. We recently reported on two new human pseudogenes of the *Adhesion* family, both similar to GPR116 [21] and these have now received official names from HUGOs Gene Nomenclature Committee and are now referred to as GPR116P1 and GPR116P2, respectively. The rodent and human *Taste2* families (Fig. 2e) have a low proportion of one-to-one orthologous pairs. Among the 35 rat, 35 mouse and 25 human receptors, there are only 10 orthologous triplets (one copy from each species). These represent 20% of all members, whereas in rats and mice, 91% of the *Taste2s* make up one-to-one orthologous pairs. The number of *Taste2s* in our datasets is equal or close to previously reported numbers [30-35]. The rat receptors T2R39 and T2R45 [33] are not available in any public sequence database.

The Vomeronasal 1 receptor family

Five full-length *Vomeronasal1* genes, VN1R1-5, have been previously identified from sequence mining in the human genome assembly [8]. We removed VN1R3 from this list because it contains a frame shift in the current genome assembly, but we found a new intact human *Vomeronasal1*, FKSG83. The number of human *Vomeronasal1* pseudogenes was first estimated at 195 [8], but this number was later reduced to 115 [9] and in this study we found only 53. We found 145 full-length genes and 164 *Vomeronasal1* pseudogenes in the mouse genome and 105 full-length genes and 84 pseudogenes in the rat genome. These numbers are within the ranges of previously reported figures which are 137-187 full-length and 156-168 pseudogenes for mouse and 95-106 full-length and 21-110 pseudogenes for rat [9-11,36,37]. We found many non-*Vomeronasal1s* and duplicates within the datasets from Young *et. al.* [9]. A phylogenetic tree of the *Vomeronasal1* family was produced [see Additional File 5]. Only 9% of the rat-mouse receptors are one-to-one orthologues and there are no human-rat one-to-one orthologues.

The vomeronasal type 2 receptors (V2Rs)

The human V2R repertoire has not been mined in our previous whole genome GPCR repertoire articles and here we identified 1 full-length gene and 11 pseudogenes. Moreover, we almost doubled the number of reported intact V2Rs in rat (from 61 to 108) and mouse (from 57 to 111)

and raised the total numbers of receptors considerably (168 to 197 in rat and from 209 to 295 in mouse) [38]. Our phylogenetic tree of the V2Rs is available as an additional file [see Additional File 6]. There are no human-rat orthologues in this receptor group and only 2% of the rodent V2Rs are orthologues.

The olfactory receptors (ORs)

The olfactory receptors make up half or more of all GPCRs in mammalian species (human: 49%, mouse 61% and rat 66%) and they constitute 1,7%, 4,5% and 5,3% of the genes in the human, mouse and rat genomes, respectively (gene numbers from Ensembl [39]). Our search for human ORs identified all (388) of the full-length genes and 96% of the pseudogenes in the HORDE database [40], and in addition we found 12 new pseudogenes. This shows that the sensitivity and specificity of our searches was high. We identified 1081 full-length OR genes and 325 OR pseudogenes in the mouse genome which is close to what was reported in a recent analysis which resulted in 1037 full-length genes and 354 pseudogenes [41]. We found 1814 ORs (1234 full-length genes, 552 pseudogenes and 28 partial genes) in the rat genome. The rat therefore has at least 153 (14%) more full-length OR genes and 227 (70%) more OR pseudogenes than mice and 846 (218%) more full-length OR genes and 73 (15%) more OR pseudogenes than humans. The percentages of pseudogenes are 23% in mouse, 31% in rats and 55% in humans. An OR phylogenetic tree with pie charts of the proportions of orthologues and the large tree file in Newick format were produced [see Additional files 7 and 8]. 8% (120) of the rat and human ORs are orthologous, 18% (268) are human specific and 74% (1115) are rat specific. Of the mouse and rat ORs 31% (550) are orthologous, 30% (534) are mouse specific and 38% (675) are rat specific.

"Other GPCRs"

The additional sequences suggested by various sources to be GPCRs, here called "Other GPCRs" or just "Others", are shown in Table 2. There are 4 pairs (GPR107–GPR108, GPR177–GPR178, GPR172A–GPR172B, TMEM185A–Tmem185a) and 2 triplets (GPR137–GPR137B–GPR137C, PAQR5–PAQR7–PAQR8) of homologues within this set. No additional sequences related to these have been found in the human, mouse and rat genomes except for the PAQRs that have additional homologues, but these are not suggested to be GPCRs [42]. Thus the "Others" (except the PAQRs) seem to make up protein families of their own comprising only 1–3 members. GPR143 (OA1) is the only one in this list that has been shown to be able to activate a G protein [43]. For PAQR5 different reports describe either plasma membrane or intracellular location [44,45]. PAQR5 GPR149 (IEDA), Gpr181p (pseudogene in human) and GPR157 have vague and ambiguous

sequence similarities to several different GPCRs families. GPR107, GPR108, GPR137, GPR137B and GPR137C show low similarity to various membrane bound proteins. Only GPR143, GPR149, Gpr181p and GPR157 show a higher similarity to membrane proteins than other proteins and 10 *Others* are indicated to have more or less than 7 transmembrane regions. Some sequences suggested to be GPCRs have later been shown to be members of other membrane protein families. We found one such example when searching in the non-redundant database for alternative versions of putative GPCR sequences. A protein [GenBank: [BAB15283.1](#)] recently discovered by searches utilising Hidden Markov Models and suggested to be a GPCR [46] is a sugar transporter, MYL5/MFSD7 [GenBank: [AAQ88767](#)] comprising 12 transmembrane helices. Two additional human GPCR sequences from the same study have now received GPR names, GPR177 and GPR178, by HGNC upon our request and are found among the *Others* in Table 2.

Discussion

Here we present the first overall analysis of the GPCR subset of the rat genome and we provide updated versions of the human and mouse GPCR repertoires. 1276 (68%) of the full-length rat GPCRs had not been previously annotated and we also identified 40 new mouse receptors and 1 new human receptor. Of the new receptors in this study, 93% are ORs and 6% are vomeronasal receptors. We have performed comprehensive sequence mining and our datasets have been carefully compared with previously available versions. The GPCR gene sequences have been verified for a complete and non-interrupted coding region and incorrectly predicted sequences have been manually curated. We also performed phylogenetic analyses displaying for the first time the detailed orthologous relationships of the GPCR repertoires in rats, mice and humans. This information is valuable when results from pharmacological and physiological studies are compared or extrapolated between species.

The authors of the draft rat genome assembly estimated the proportion of one-to-one orthologues to be 89–90% between the rat and human genomes and 86–94% between the rat and mouse genomes [1]. The average amino acid sequence identities of the orthologues were reported to be 88.3% for rat and human and 95.0% for rat and mouse. Remarkably, we find that the average proportion of one-to-one orthologues within the GPCR superfamily is much lower. The average proportions of such orthologues were only 58% for the rat and human, and 70% for the rat and mouse GPCR repertoires. The average protein sequence identities of the GPCR orthologue pairs is also lower than for the whole genomes. We found these to be 80% for the rat and human pairs and 90% for the rat and mouse pairs. This suggests that the GPCR superfamily

Table 2: "Other GPCRs"

GPCR (alias)	#TM helices	Conserved domains (e-value)	First BLAST hits (best e-value)
PAQR5	6	HlyIII (3e-23)	PAQR8 (6e-26), PAQR7 (6e-24), other PAQRs
PAQR7	6	HlyIII (2e-20)	PAQR8 (2e-73), PAQR5 (1e-25), other PAQRs
PAQR8	6	HlyIII (4e-27)	PAQR7 (6e-24), PAQR5 (1e-26), other PAQRs
GPR143 (OAI)	6	7TM_2 (3e-3), 7TM_1 (0.017)	Secretin (0.09), Adhesion (0.091)
GPR149 (IEDA)	7	7TM_1 (1e-6)	Rhodopsin (0.5)
Gpr181p	7	7TM_1 (2e-3)	VIR (3e-5), Rhodopsin (3e-4), TAS2R (9e-3)
GPR157	7	7TM_2 (6e-13), 7TM_1 (8e-6), FZD (2e-4)	Secretin (1e-4), Rhodopsin (1e-4), Adhesion (9e-4)
GPR107	7	Lung_7-TM_R (9e-76)	GPR108 (2e-130), membrane proteins (1e-6)
GPR108	7	Lung_7-TM_R (9e-78)	GPR107 (e-132), membrane proteins (4e-8)
GPR137 (C11ORF4)	7	-	GPR137B (9e-75), membrane proteins (7e-16)
GPR137B (TM7SF1)	7	-	GPR137 (6e-99), membrane proteins (1e-6)
GPR137C (TM7SF1L2)	7	-	GPR137B (3e-60), different proteins (1e-4)
TM7SF3	7	-	-
TM7SF4 (FIND, DC-STAMP)	6	-	different proteins (3e-9)
GPR175	7	-	-
GPR177	8	DUF1171 (4e-130)	GPR178 (6e-5)
GPR178	9	MARCKS (1e-3), STOP (5e-3), APC (7e-3), IER (8e-3)	GPR177 (1e-5)
TMEM185A	8	-	Tmem185b (4e-169), uncharacterised proteins
Tmem185b	7	-	TMEM185A (7e-169), uncharacterised proteins
GPR172A (PERVAR1)	11	DUF1011 (3e-29)	GPR172B (4e-136), uncharacterised proteins
GPR172B (PERVAR2)	11	DUF1011 (3e-32)	GPR172A (1e-144), uncharacterised proteins

"Other GPCRs"; sequences that have been suggested to be GPCRs but are not members of any GPCR family. The searches for transmembrane helices were made using TMHMM version 2.0 [86]. Conserved domains were searched for using conserved domain search [71] with a cut-off at $e = 0.01$. The BLAST searches for the most similar proteins were performed in the Refseq database using a cut-off at $e = 0.01$.

is much more divergent than the protein families in the genome in general. However, this difference in the proportions and sequence identities of orthologues may not entirely be explained by the relatively high divergence of the GPCR superfamily. This is also related to the fact that we have mined and edited the sequences to a much more complete level than the sequence data from the original draft rat genome assembly.

There are very different levels of evolutionary conservation or orthologous pairing of the GPCR families. There are families in which all (*Glutamate* (except the V2R group) and *Secretin* families) or almost all (*Adhesion* and *Frizzled*) receptors are conserved between the compared species. Other families have expanded separately in the three species (primarily the *Taste2s*) or expanded in one lineage while vanishing (*Vomeronsal1* and V2R) in another. There are also remarkable examples of family subgroups that have all expanded in the rodents while diminishing in primates (olfactory receptors (ORs), trace-amine-associated receptors (TAARs), MAS-related GPCRs (MRGs), formyl peptide receptor like receptors (FPRLs) and vomeronasal type 2 receptors (V2Rs)). It is notable that all the large inter-species variations are found among the families and subgroups that respond to exogenous stimuli. Thus these differences are likely to make important contributions to the diversification of the senses of

olfaction, taste and pheromone sensation among the species. The pheromone sensation has evolved in rats and mice, which have many species-specific vomeronasal receptors (*Vomeronsal1s* and V2Rs), but disappeared in the primate lineage in which the vomeronasal organ is regressed in adults and critical components of vomeronasal transduction pathway have been lost [47,48]. Moreover, it has recently been shown that the TAARs may act as chemosensory receptors in the olfactory epithelium [49,50].

There is a considerable variation among the different families and groups of GPCRs in the percentage of amino acid identity of orthologues. The rat and human *Frizzled* family orthologues have 95% average sequence identity while the lowest percentage identity between rats and human orthologous pairs is found in the *Taste2* family (58%). The *Taste2* family also display a very high divergence with respect to the repertoire. Only 20% of the rat and human and 84% of the rat and mouse receptors are one-to-one orthologues, indicating that this family is evolving rapidly. The *Taste2* receptors mediate the sense of bitter taste [51] and it is likely that their divergence has contributed to the different capacities of species to recognise bitter tastants. Also the *Adhesion* family GPCRs display relatively low amino acid identity between rat, mouse and human orthologues (72%). However, in contrast to the *Taste2*

family, the *Adhesion* family repertoire is relatively well conserved; 100% of the rat and mouse and 91% of the rat and human *Adhesions* make up one-to-one orthologous pairs. The fact that the *Adhesions* have been retained during evolution shows that they have important physiological functions and their relatively low sequence conservation indicate that their action is less dependent on the transmembrane region. This is in agreement with the hypothesis that the transmembrane regions of the *Adhesion* GPCRs may perhaps not be involved in direct interactions with ligands or G-proteins but have an important role as membrane anchors [52].

The human ORs have previously been divided into class I and class II based on phylogenetic criteria [53]. The human genome contains 53 (14%) class I and 335 (86%) class II ORs and our analysis shows that the rat genome includes 139 (11%) class I ORs and 1096 (89%) class II. Class I ORs mediate the effect of water-soluble odorants and class II receptors recognise airborne odorants [53,54]. The proportion of rat and mouse OR orthologues is relatively low (31%) and the proportion of rat and human OR orthologues is very low (8%). Species-specific receptors can be found in most parts of the phylogenetic tree, but there are also several large clusters of species-specific ORs [see Additional File 8]. There are five large clusters containing exclusively rat-specific ORs. These are located closest to the human OR1F (18 rat ORs), OR111 (7 rat ORs), OR7A5,10,17 (6 rat ORs), OR2AK2 (6 rat ORs), OR2B11 (9 rat ORs) (for nomenclature, see [53]). Many human-specific ORs were found in the OR2T-group which contains 13 human ORs (OR2T1-7, 10-11, 27, 29, 34, 35) without a rodent orthologue. These species-specific clusters include receptors exclusively belonging to class II suggesting that this differentiation between the species has contributed to the unique capacities to recognise volatile airborne substances.

It is difficult to envision why in some cases several species lack a certain receptor while the same receptor has been retained during evolution in other lineages/species. One of the most likely explanations is that the gene has been lost because another gene has taken over its functions. One example of this could be the motilin receptor, MLNR. We did not find orthologues to this human receptor in neither rat nor mouse. In further searches (data not shown) we did however, find orthologues for this receptor in chimps (partial sequence), cow, chicken, zebrafish and pufferfish (*Tetraodon nigroviridis*) but not in dogs, suggesting that this receptor has been specifically lost in several mammals. MLNR is most closely related to the ghrelin receptor, GHSR (named after the subfragment of the growth hormone secretagogue peptide) and these receptors have 44.2% overall protein sequence identity. Motilin stimulates gastrointestinal motility and this function is

modulated by a pharmaceutical substance, erythromycin A, which mimics this peptide and targets the MLNR receptor [55,56]. It has been shown that the ghrelin peptide with its receptor regulates gastrointestinal motility in rodents and the expression of GHSR in the human gastrointestinal tract indicates that ghrelin plays a role in gastrointestinal motility in humans as well [57-59]. It is thus possible that the ghrelin system plays a compensating role in GI motility in species that are missing the motilin receptor, such as rodents and dogs.

There are several additional cases in which two *Rhodopsin* family receptors have the same ligand and could share functions in some species while one of the receptors has been lost in several other evolutionary lineages such that the other gene may have taken over. TRHR and Trhr2 both bind to thyrotropin-releasing hormone (TRH). The former receptor has been conserved throughout all searched vertebrate genomes whereas Trhr2 is lacking in primates, chicken and fugu suggesting that it has been lost independently in three lineages. Furthermore the neuropeptide B/W receptors (NPBWR1 and NPBWR2) have a 62.2% overall sequence identity in human. We found both NPBWR1 and NPBWR2 in fish but whereas NPBWR1 has been retained in the examined vertebrate species NPBWR2 is found neither in the dog nor the chicken genomes and has become a pseudogene in the rodent lineage. Another similar case is seen for the melanin-concentrating hormone receptors MCHR1 and MCHR2. We could find both receptor subtypes in fish but whereas the former has been conserved the latter, MCHR2, is absent or is a pseudogene in rat, mouse, hamster, guinea pig and chicken [60]. This is illustrated also for another *Rhodopsin* family receptor, OXER1, which is missing in the rodents and in chicken but is conserved in fish, primates, cow and dog (1 full-length gene and 1 pseudogene). For this receptor there exists no characterized homologue that binds the same ligand (5-oxo-ETE) but there are several closely related orphan receptors such as GPR31, GPR81, GPR109A and GPR109B (primate specific duplicate) that could tentatively be activated by the same ligand.

In summary, we have presented the overall GPCR repertoire of rats and analysed it in relation to updated datasets for human and mouse. We have identified new genes, made extensive comparisons to previously published datasets and performed phylogenetic analyses to distinguish orthologous and species-specific receptors. This material is important for the interpretation of studies in which results are extrapolated between species to provide information about receptor structure or native ligand.

Conclusion

We present the first overall analysis of the GPCR repertoire in rats and compare this to versions of those in mouse and human. The receptor sequences have been manually curated to assure a higher level of completeness and quality. The detailed relationships of the repertoires, here determined by careful phylogenetic analyses, are valuable information when results from pharmacological and physiological studies are compared or extrapolated between species. The greatest diversification, with regard to both receptor numbers and sequences, is seen for those GPCRs that respond to exogenous stimuli indicating that the alteration of the repertoires of these receptors has been important for adaptation of the species to their environment. Our detailed comparison of GPCR repertoires also revealed several examples of species-specific expansions and deletions. We performed an extended analysis in additional species of several *Rhodopsins* that were shown to have been lost in different lineages but retained in others, investigating if their function might be shared with another (homologous) receptor. Most *Rhodopsin* family GPCRs that bind the same type of ligands, e.g. lipids and peptides, formed coherent groups in our phylogenetic analysis. However, many (28) of the human orphan *Rhodopsins* could not be grouped with any characterized receptor because of low sequence identity and/or ambiguous relationships making the prediction of the ligand difficult. We found the number of human orphan *Rhodopsin* receptors to be 65 while the number was 84 in 2002 [61] suggesting that the current rate of de-orphanization is higher than that for the discovery of members of this family.

Methods

Identification of rat GPCRs of the Glutamate, Rhodopsin, Adhesion, Frizzled and Secretin family receptors

The mouse GPCRs of the *Glutamate* (except V2R group), *Rhodopsin* (except OR group), *Adhesion*, *Frizzled* and *Secretin* families [21] were used as baits in stand alone BLASTP and TBLASTN searches [62] against NCBI non-redundant (nr) database [63]. For each and every query the accession numbers of the 20 first hits were collected and a non-redundant list was obtained which was used to collect the sequences of the hits. Sequence duplicates with different names were identified using BLASTCLUST [62] with the threshold set at 99% sequence identity. Remaining duplicates were splice variants or contained sequence polymorphisms or sequencing errors exceeding a base difference of 1% of the total sequence. These were identified from their genomic overlap as determined using stand alone BLAT against the June 2003 rat genome assembly. Missing rat orthologues were searched for using online BLAT [64] and TBLASTN against the nr and Celera databases [65].

Identification of the human, rat and mouse OR, Taste2, Vomeronasal1 and V2R receptors

TBLASTN searches in genome assemblies

Query datasets were retrieved by downloading a start set of human, rat and mouse GPCR protein sequences from the Genbank database using the Entrez data-retrieval tool [66] in keyword searches. The downloaded datasets were searched using TBLASTN against the October 2005 releases of the Ensembl human (build 35), mouse (build 34) and rat (assembly 3.4) genome assemblies [5]. For the OR, *Taste2* and *Vomeronasal1* receptors, which are all intronless, we used the whole protein sequences. For the V2R, which have several exons, we used only the transmembrane region which is located within one exon. The cut offs used were $e = 0.01$ for the OR and rat V2R searches and $e = 1$ for the *Taste2* and *Vomeronasal1* searches.

Extracting gene sequences and removing non-GPCRs

The TBLASTN searches result lists, previously saved in tabular format, were processed so that overlapping hits on the same strand were merged using a custom made Java program (available upon request). The chromosome coordinates of the merged hits were used to extract the corresponding sequences from the genome using `fastacmd` of the NCBI blast package. Each sequence was elongated upstream until the first start codon and downstream until the first stop codon using a custom made Java program (available upon request). The preliminary datasets were "cleaned" from non-GPCRs and GPCRs from other families by searching it against the Refseq database using BLASTN with the default settings. The criterion used when including new GPCRs was that at least the first two hits had to belong to the same GPCR family. Protein translations were obtained using `transeq` from the EMBOSS package [67] and from these the longest intact coding domain was extracted. Protein translations were considered full-length proteins if they contained an intact seven transmembrane region. Proteins containing 5 or more unknown positions due to incompleteness in the genome assemblies were removed and put into a "partial" category of gene sequences.

Naming the GPCRs

The human receptors have been named using the official Gene name and rat and mouse orthologues have been named after their human counterpart. For those human GPCRs which lacked official Gene names we requested new names from the HUGO Gene Nomenclature Committee (HGNC) [68]. We obtained new names for GPR177, GPR178 and GPR181P. We also received the official names for GPR116P1, GPR116P2, GPR166P and P2RY10P2 which had been previously been named by HGNC. The human ORs were named according to a widely accepted nomenclature (e.g. OR1A2) after matching our dataset to that of the HORDE database [40]. The

Vomerinasal1s were named after existing nomenclatures for human [8], rat [11] and mouse [37,69]. There existed a nomenclature for the rodent V2Rs [38], which we used. For the *Taste2s* we used the official Gene names which are T2R-X in human and Tas2r-X in mouse, where X is a number. In rat these two nomenclatures are mixed and several members lack a Gene name. Other new receptors were assigned labels according to their chromosome coordinates prefixed by a letter indicating the species: human (H), rat (r) and mouse (m).

Phylogenetic analyses

Removing non-conserved sequence regions

For the *Rhodopsins* the N-termini, C-termini and loops were defined from alignments to the protein sequence of the crystallised bovine Rhodopsin structure [70] and removed. For the other families NCBI's conserved domain search [71] was used to identify the borders of the transmembrane region and removing the N- and C-termini. The ORs and *Vomerinasal1s* were cut according to the "7tm_1" domain (*Rhodopsin* family), the *Taste2s* to the TAS2R domain, the *Glutamate* receptors (including the V2Rs) to the "7tm_3" domain (metabotropic glutamate family) and the *Adhesions* and *Secretins* to the "7tm_2" domain (Secretin family).

Division of the Rhodopsin family dataset

BLASTCLUST was used to apply a minimum threshold of protein sequence identity on a dataset consisting of all human *Rhodopsins*. The receptor with the lowest identity (less than 23%) to other *Rhodopsins* is GPR160 and this receptor was the first to be separated from the remaining dataset (Fig. 2). The clustering and separation procedure was repeated at increasing thresholds of sequence identity allowing the iterative separation of groups, pairs and single members with atypical sequences. This division of the dataset continued until we obtained reproducible topology of the human *Rhodopsins* in neighbor joining, maximum parsimony and maximum likelihood analyses (data not shown). This was achieved after the 36% threshold for sequence identity. All sequences that were separated from the main dataset were subject to BLAST searches against the whole human dataset. Queried/searched sequences which had all first five BLAST hits in the same cluster of related sequences were considered distant relatives and were merged with that particular cluster/dataset. The rat *Rhodopsin* receptors were merged with the corresponding dataset of human orthologues. A length threshold of 50% was used for the BLASTCLUST clustering.

Sequence aligning and bootstrapping

All datasets were aligned using CLUSTALW 1.82 [72] with default alignment parameters. Pseudogenes were excluded and the transmembrane regions extracted as described above. The stability of the phylogenetic branch-

ing was determined using bootstrap analyses. The alignments were bootstrapped 100 times using SEQBOOT from the UNIX version of the PHYLIP 3.6 package [73], except for the olfactory receptor (OR) alignment which was bootstrapped 10 times. This resulted in a total of 100 (10 for the ORs) different alignments from the original dataset, respectively.

Calculating phylogenetic trees

Unless otherwise specified the bootstrapped files were used for calculating maximum parsimony (MP) trees with PROTPARS from the PHYLIP 3.6 package. The trees were un-rooted and calculated using ordinary parsimony and the topologies were obtained using the built-in tree search procedure. From the bootstrapped OR alignment files protein distances were calculated using PROTDIST from the Win32 version of the PHYLIP 3.6 package. The Jones-Taylor-Thornton matrix was used for the calculation. The trees were calculated from the 10 different distance matrices, previously generated with PROTDIST, using NEIGHBOR from the same package. For the *Rhodopsin* family the largest (non-OR) group (Figure 2) was analysed with both MP and neighbor joining (NJ), whereas maximum likelihood (ML) was used for all other groups (Figure 3). For the latter groups branch lengths were calculated using the ML consensus tree as a user defined tree in TreePuzzle [74]. The following parameters were used; Type of analysis: Tree reconstruction; Tree-search procedure: User-defined trees; Compute clocklike branch lengths: Yes; Location of root: Best Place (automatic search); Parameter estimates: Exact (slow); Parameter-estimation uses: 1st input tree; Type of sequence input data: Amino acids; Model of substitution: VT Mueller-Vingron Model of Substitution, 2000; Amino acid frequencies: Estimate from dataset; Model of rate heterogeneity: Mixed (one invariable plus eight Gamma rates); Fraction of invariable sites: Estimate from dataset; Gamma distribution parameter alpha: Estimate from dataset; Number of Gamma rate categories: eight.

Merging and viewing the trees

The tree files were merged using the GNU-UNIX cat command and the resulting files were analysed using CONSENSE, from the PHYLIP 3.6 package, to retrieve bootstrapped consensus trees. The largest human-rat *Rhodopsin* family tree is a consensus tree of two other consensus trees, one from the NJ analysis and the other from the MP analysis. The trees were plotted using TREEVIEW and manually edited in CANVAS.

Ligand type classification for the human Rhodopsins

An in-house list of ligand types for the human *Rhodopsins* was updated by obtaining missing ligand information from the receptor list maintained by the International Union of Pharmacology Committee on Receptor Nomen-

clature and Drug Classification (NC-IUPHAR) [75]. Receptors that did not have a ligand in this list were considered candidate orphan receptors. These were systematically searched using keyword look-ups in Pubmed and BLAST in the NCBI's non-redundant (nr) database to find receptor deorphanization reports [26,76-85]. The same GPCR often has several names and in order to obtain a higher coverage of publications we used names found from HGNC, NCBI Gene and the identical BLAST hits.

Abbreviations

GPCR, G protein-coupled receptor; *Taste2*, taste receptor type 2 family of GPCRs; *Vomeronasal1*, vomeronasal receptor type 1 family of GPCRs; *Others* Other GPCRs; OR, olfactory receptor; V2R, vomeronasal type 2 receptor; HGNC, HUGO Gene Nomenclature Committee

Competing interests

The author(s) declares that there are no competing interests.

Authors' contributions

DEG carried out the receptor identification, performed the phylogenetic analyses and drafted the manuscript. RF participated in the design of the study and helped to draft the manuscript. HBS conceived the study, participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Comparative table of the GRAFS(T) family receptors in human, rat and mouse. A table listing the rat, mouse and human GPCRs of the Glutamate, Rhodopsin, Adhesion, Frizzled, Secretin and Taste2 families. Pseudogenes are marked with a "P" in the column to the right of the respective species. When several species- or lineage-specific duplicates (paralogues) exist the paralogue with the highest sequence identity has been given as the primary orthologue whereas the other genes are present on separate rows in the table and without a counterpart in the other species.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-8-338-S1.xls>]

Additional file 2

Rat GPCR gene sequences. The nucleotide sequences of all rat GPCRs compiled in this analysis, including pseudogenes. Note that this is a large file which will cover hundreds of pages if printed.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-8-338-S2.txt>]

Additional file 3

Human GPCR gene sequences. The nucleotide sequences of all human GPCRs compiled in this analysis, including pseudogenes. Note that this is a large file which will cover hundreds of pages if printed.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-8-338-S3.txt>]

Additional file 4

Mouse GPCR gene sequences. The nucleotide sequences of all mouse GPCRs compiled in this analysis, including pseudogenes. Note that this is a large file which will cover hundreds of pages if printed.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-8-338-S4.txt>]

Additional file 5

Phylogenetic tree of the Vomeronasal1 family. The figure shows the consensus trees of 100 maximum parsimony trees of the human and rat Vomeronasal1 GPCR family. The first pie chart above the tree shows the proportions of human-rat one-to-one orthologues (O), human specific (H) and rat specific (R) members. The second pie chart displays the proportions of rat-mouse one-to-one orthologues (O), rat specific (R) and mouse specific (M) members.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-8-338-S5.pdf>]

Additional file 6

Phylogenetic tree of the vomeronasal type 2 receptors (V2Rs). The figure shows the consensus trees of 100 maximum parsimony trees of the human and rat vomeronasal type 2 receptors (V2Rs). The first pie chart above the tree shows the proportions of human-rat one-to-one orthologues (O), human specific (H) and rat specific (R) members. The second pie chart from the top displays the proportions of rat-mouse one-to-one orthologues (O), rat specific (R) and mouse specific (M) members.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-8-338-S6.pdf>]

Additional file 7

Phylogenetic tree of the olfactory receptors (ORs). The figure shows the consensus tree of 10 neighbor joining trees of the human, rat and mouse olfactory receptors. The pie charts to the left show the proportions of human-rat one-to-one orthologues (O), human specific (H) and rat specific (R) members. The pie charts to the right display the proportions of rat-mouse one-to-one orthologues (O), rat specific (R) and mouse specific (M) members. The pie charts on the top give the proportions of one-to-one orthologues for all ORs, whereas the charts below contain the same information for the Class I and Class II subsets, respectively.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-8-338-S7.pdf>]

Additional file 8

Phylogenetic tree file of the olfactory receptors (ORs). A Newick format consensus tree file from 10 neighbor joining trees of the human, rat and mouse olfactory receptors. Note that this is a very large tree file which can be difficult to open with many tree viewing programs and will cover a very large number of pages if printed. The file can be viewed in most tree viewing programs for example TREEVIEW which is available at no cost.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-8-338-S8.tre>]

Acknowledgements

The studies were supported by the Swedish Research Council (VR, medicine), the Swedish Brain Foundation, Svenska Läkaresällskapet, Åke Wiberg Foundation, Lars Hiertas foundation, Thuring's foundation, The Novo Nordisk Foundation, and the Magnus Bergwall Foundation.

References

- Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, Okwuonu G, Hines S, Lewis L, DeRamo C, Delgado O, Dugan-Rocha S, Miner G, Morgan M, Hawes A, Gill R, Celera, Holt RA, Adams MD, Amanatides PG, Baden-Tillson H, Barnstead M, Chin S, Evans CA, Ferriera S, Fosler C, Glodek A, Gu Z, Jennings D, Kraft CL, Nguyen T, Pfannkoch CM, Sitter C, Sutton GG, Venter JC, Woodage T, Smith D, Lee HM, Gustafson E, Cahill P, Kana A, Doucette-Stamm L, Weinstock K, Fechtel K, Weiss RB, Dunn DM, Green ED, Blakesley RW, Bouffard GG, De Jong PJ, Osoegawa K, Zhu B, Marra M, Schein J, Bosdet I, Fjell C, Jones S, Krzywinski M, Mathewson C, Siddiqui A, Wye N, McPherson J, Zhao S, Fraser CM, Shetty J, Shatsman S, Geer K, Chen Y, Abramzon S, Nierman WC, Havlak PH, Chen R, Durbin KJ, Egan A, Ren Y, Song XZ, Li B, Liu Y, Qin X, Cawley S, Cooney AJ, D'Souza LM, Martin K, Wu JQ, Gonzalez-Garay ML, Jackson AR, Kalafatis KJ, McLeod MP, Milosavljevic A, Virk D, Volkov A, Wheeler DA, Zhang Z, Bailey JA, Eichler EE, Tuzun E, Birney E, Mongin E, Ureta-Vidal A, Woodwark C, Zdobnov E, Bork P, Suyama M, Torrents D, Alexandersson M, Trask BJ, Young JM, Huang H, Wang H, Xing H, Daniels S, Gietzen D, Schmidt J, Stevens K, Vitt U, Wingrove J, Camara F, Mar Alba M, Abril JF, Guigo R, Smit A, Dubchak I, Rubin EM, Couronne O, Poliakov A, Hubner N, Ganten D, Goesele C, Hummel O, Kreitler T, Lee YA, Monti J, Schulz H, Zimdahl H, Himmelbauer H, Lehrach H, Jacob HJ, Bromberg S, Gullings-Handley J, Jensen-Seaman MI, Kwitek AE, Lazar J, Pasko D, Tonellato PJ, Twigger S, Ponting CP, Duarte JM, Rice S, Goodstadt L, Beaton SA, Ames RD, Winter EE, Webber C, Brandt P, Nyakatura G, Adetobi M, Chiaromonte F, Elnitski L, Eswara P, Hardison RC, Hou M, Kolbe D, Makova K, Miller W, Nekrutenko A, Riemer C, Schwartz S, Taylor J, Yang S, Zhang Y, Lindpaintner K, Andrews TD, Caccamo M, Clamp M, Clarke L, Curwen V, Durbin R, Eyrae E, Searle SM, Cooper GM, Batzoglou S, Brudno M, Sidow A, Stone EA, Payseur BA, Bourque G, Lopez-Otin C, Puente XS, Chakrabarti K, Chatterji S, Dewey C, Pachter L, Bray N, Yap VB, Caspi A, Tesler G, Pevzner PA, Haussler D, Roskin KM, Baertsch R, Clawson H, Furey TS, Hinrichs AS, Karolchik D, Kent WJ, Rosenbloom KR, Trumbower H, Weirauch M, Cooper DN, Stenson PD, Ma B, Brent M, Arumugam M, Shteynberg D, Copley RR, Taylor MS, Riethman H, Mudunuri U, Peterson J, Guyer M, Felsenfeld A, Old S, Mockrin S, Collins F: **Genome sequence of the Brown Norway rat yields insights into mammalian evolution.** *Nature* 2004, **428**:493-521.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Showkeen R, Sims S, Waterston RH, Wilson RK, Hillier LV, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickinson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Mimosima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramsay J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korfi I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrino A, Morgan MJ, Szustakowski J, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusser D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson C, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams W, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T,

- Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X: **The sequence of the human genome.** *Science* 2001, **291**:1304-1351.
4. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Graffham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery K, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Raymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapozhnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420**:520-562.
 5. **The Ensembl FTP site** [<http://ftp.ensembl.org/pub>]
 6. Rogic S, Mackworth AK, Ouellette FB: **Evaluation of gene-finding programs on mammalian sequences.** *Genome Res* 2001, **11**:817-832.
 7. Lagerstrom MC, Hellstrom AR, Gloriam DE, Larsson TP, Schiöth HB, Fredriksson R: **The G Protein-Coupled Receptor Subset of the Chicken Genome.** *PLoS Comput Biol* 2006, **2**:e54.
 8. Rodriguez I, Mombaerts P: **Novel human vomeronasal receptor-like genes reveal species-specific families.** *Curr Biol* 2002, **12**:R409-11.
 9. Young JM, Kambere M, Trask BJ, Lane RP: **Divergent VIR repertoires in five species: Amplification in rodents, decimation in primates, and a surprisingly small repertoire in dogs.** *Genome Res* 2005, **15**:231-240.
 10. Shi P, Bielawski JP, Yang H, Zhang YP: **Adaptive diversification of vomeronasal receptor I genes in rodents.** *J Mol Evol* 2005, **60**:566-576.
 11. Grus WE, Zhang J: **Rapid turnover and species-specificity of vomeronasal pheromone receptor genes in mice and rats.** *Gene* 2004, **340**:303-312.
 12. Bockaert J, Pin JP: **Molecular tinkering of G protein-coupled receptors: an evolutionary success.** *Embo J* 1999, **18**:1723-1729.
 13. Birnbaumer L, Abramowitz J, Brown AM: **Receptor-effector coupling by G proteins.** *Biochim Biophys Acta* 1990, **1031**:163-224.
 14. Drews J: **Drug discovery: a historical perspective.** *Science* 2000, **287**:1960-1964.
 15. Hopkins AL, Groom CR: **The druggable genome.** *Nat Rev Drug Discov* 2002, **1**:727-730.
 16. Fredriksson R, Lagerstrom MC, Lundin LG, Schiöth HB: **The g-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints.** *Mol Pharmacol* 2003, **63**:1256-1272.
 17. Fredriksson R, Schiöth HB: **The repertoire of G-protein-coupled receptors in fully sequenced genomes.** *Mol Pharmacol* 2005, **67**:1414-1425.
 18. Attwood TK, Findlay JB: **Fingerprinting G-protein-coupled receptors.** *Protein Eng* 1994, **7**:195-203.
 19. Kop EN, Kwakkenbos MJ, Teske GJ, Kraan MC, Smeets TJ, Stacey M, Lin HH, Tak PP, Hamann J: **Identification of the epidermal growth factor-TM7 receptor EMR2 and its ligand dermatan sulfate in rheumatoid synovial tissue.** *Arthritis Rheum* 2005, **52**:442-450.
 20. Kwakkenbos MJ, Pouwels W, Matmati M, Stacey M, Lin HH, Gordon S, van Lier RA, Hamann J: **Expression of the largest CD97 and EMR2 isoforms on leukocytes facilitates a specific interaction with chondroitin sulfate on B cells.** *J Leukoc Biol* 2005, **77**:112-119.
 21. Bjarnadottir TK, Gloriam DE, Hellstrand SH, Kristiansson H, Fredriksson R, Schiöth HB: **Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse.** *Genomics* 2006.
 22. Vassilatis DK, Hohmann JG, Zeng H, Li F, Ranchalis JE, Mortrud MT, Brown A, Rodriguez SS, Weller JR, Wright AC, Bergmann JE, Gaitanaris GA: **The G protein-coupled receptor repertoires of human and mouse.** *Proc Natl Acad Sci U S A* 2003, **100**:4903-4908.
 23. Dong X, Han S, Zylka MJ, Simon MI, Anderson DJ: **A diverse family of GPCRs expressed in specific subsets of nociceptive sensory neurons.** *Cell* 2001, **106**:619-632.
 24. Gloriam DE, Bjarnadottir TK, Yan YL, Postlethwait JH, Schiöth HB, Fredriksson R: **The repertoire of trace amine G-protein-coupled receptors: large expansion in zebrafish.** *Mol Phylogenet Evol* 2005, **35**:470-482.
 25. Lindemann L, Ebeling M, Kratochwil NA, Bunzow JR, Grandy DK, Hoener MC: **Trace amine-associated receptors form structurally and functionally distinct subfamilies of novel G protein-coupled receptors.** *Genomics* 2005, **85**:372-385.
 26. Ciana P, Fumagalli M, Trincavelli ML, Verderio C, Rosa P, Lecca D, Ferrario S, Parravicini C, Capra V, Gelosa P, Guerrini U, Belcredito S, Cimino M, Sironi L, Tremoli E, Rovati GE, Martini C, Abbracchio MP: **The orphan receptor GPR17 identified as a new dual uracil nucleotides/cysteinyl-leukotrienes receptor.** *Embo J* 2006, **25**:4615-4627.
 27. Surgand JS, Rodrigo J, Kellenberger E, Rognan D: **A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors.** *Proteins* 2006, **62**:509-538.
 28. Sexton PM, Morfis M, Tilakaratne N, Hay DL, Udawela M, Christopoulos G, Christopoulos A: **Complexing receptor pharmacology: modulation of family B G protein-coupled receptor function by RAMPs.** *Ann N Y Acad Sci* 2006, **1070**:90-104.
 29. Bjarnadottir TK, Fredriksson R, Hoglund PJ, Gloriam DE, Lagerstrom MC, Schiöth HB: **The human and mouse repertoire of the adhesion family of G-protein-coupled receptors.** *Genomics* 2004, **84**:23-33.
 30. Go Y, Satta Y, Takenaka O, Takahata N: **Lineage-specific loss of function of bitter taste receptor genes in humans and nonhuman primates.** *Genetics* 2005, **170**:313-326.
 31. Conte C, Ebeling M, Marcuz A, Nef P, Andres-Barquin PJ: **Evolutionary relationships of the Tas2r receptor gene families in mouse and human.** *Physiol Genomics* 2003, **14**:73-82.
 32. Wu SV, Chen MC, Rozengurt E: **Genomic organization, expression, and function of bitter taste receptors (T2R) in mouse and rat.** *Physiol Genomics* 2005, **22**:139-149.
 33. Shi P, Zhang J: **Contrasting modes of evolution between vertebrate sweet/umami receptor genes and bitter receptor genes.** *Mol Biol Evol* 2006, **23**:292-300.
 34. Fischer A, Gilad Y, Man O, Paabo S: **Evolution of bitter taste receptors in humans and apes.** *Mol Biol Evol* 2005, **22**:432-436.
 35. Shi P, Zhang J, Yang H, Zhang YP: **Adaptive diversification of bitter taste receptor genes in mammalian evolution.** *Mol Biol Evol* 2003, **20**:805-814.
 36. Zhang X, Rodriguez I, Mombaerts P, Firestein S: **Odorant and vomeronasal receptor genes in two mouse genome assemblies.** *Genomics* 2004, **83**:802-811.
 37. Del Punta K, Rothman A, Rodriguez I, Mombaerts P: **Sequence diversity and genomic organization of vomeronasal receptor genes in the mouse.** *Genome Res* 2000, **10**:1958-1967.
 38. Yang H, Shi P, Zhang YP, Zhang J: **Composition and evolution of the V2r vomeronasal receptor gene repertoire in mice and rats.** *Genomics* 2005, **86**:306-315.
 39. **Ensembl** [<http://www.ensembl.org>]
 40. **HORDE database** [<http://biportal.weizmann.ac.il/HORDE>]

41. Niimura Y, Nei M: **Comparative evolutionary analysis of olfactory receptor gene clusters between humans and mice.** *Gene* 2005, **346**:13-21.
42. Deckert CM, Heiker JT, Beck-Sickinger AG: **Localization of novel adiponectin receptor constructs.** *J Recept Signal Transduct Res* 2006, **26**:647-657.
43. Innamorati G, Piccirillo R, Bagnato P, Palmisano I, Schiaffino MV: **The melanosomal/lysosomal protein OAI has properties of a G protein-coupled receptor.** *Pigment Cell Res* 2006, **19**:125-135.
44. Thomas P, Pang Y, Dong J, Groenen P, Kelder J, de Vlieg J, Zhu Y, Tubbs C: **Steroid and G protein binding characteristics of the seatrout and human progesterin membrane receptor alpha subtypes and their evolutionary origins.** *Endocrinology* 2007, **148**:705-718.
45. Krietsch T, Fernandes MS, Kero J, Losel R, Heyens M, Lam EW, Huhtaniemi I, Brosens JJ, Gellersen B: **Human homologs of the putative G protein-coupled membrane progesterin receptors (mPRalpha, beta, and gamma) localize to the endoplasmic reticulum and are not activated by progesterone.** *Mol Endocrinol* 2006, **20**:3146-3164.
46. Wistrand M, Kall L, Sonnhammer EL: **A general model of G protein-coupled receptor sequences and its application to detect remote homologs.** *Protein Sci* 2006, **15**:509-521.
47. Witt M, Wozniak W: **Structure and function of the vomeronasal organ.** *Adv Otorhinolaryngol* 2006, **63**:70-83.
48. Liman ER: **Use it or lose it: molecular evolution of sensory signaling in primates.** *Pflugers Arch* 2006, **453**:125-131.
49. Liberles SD, Buck LB: **A second class of chemosensory receptors in the olfactory epithelium.** *Nature* 2006, **442**:645-650.
50. Fleischer J, Schwarzenbacher K, Breer H: **Expression of Trace Amine-Associated Receptors in the Grueneberg Ganglion.** *Chem Senses* 2007.
51. Adler E, Hoon MA, Mueller KL, Chandrashekar J, Ryba NJ, Zuker CS: **A novel family of mammalian taste receptors.** *Cell* 2000, **100**:693-702.
52. Stacey M, Lin HH, Gordon S, McKnight AJ: **LNB-TM7, a group of seven-transmembrane proteins related to family-B G-protein-coupled receptors.** *Trends Biochem Sci* 2000, **25**:284-289.
53. Glusman G, Bahar A, Sharon D, Pilpel Y, White J, Lancet D: **The olfactory receptor gene superfamily: data mining, classification, and nomenclature.** *Mamm Genome* 2000, **11**:1016-1023.
54. Freitag J, Krieger J, Strotmann J, Breer H: **Two classes of olfactory receptors in *Xenopus laevis*.** *Neuron* 1995, **15**:1383-1392.
55. Kondo Y, Torii K, Itoh Z, Omura S: **Erythromycin and its derivatives with motilin-like biological activities inhibit the specific binding of [125I]-motilin to duodenal muscle.** *Biochem Biophys Res Commun* 1988, **150**:877-882.
56. Peeters T, Matthijs G, Depoortere I, Cachet T, Hoogmartens J, Vantrappen G: **Erythromycin is a motilin receptor agonist.** *Am J Physiol* 1989, **257**:G470-4.
57. Masuda Y, Tanaka T, Inomata N, Ohnuma N, Tanaka S, Itoh Z, Hosoda H, Kojima M, Kangawa K: **Ghrelin stimulates gastric acid secretion and motility in rats.** *Biochem Biophys Res Commun* 2000, **276**:905-908.
58. Takeshita E, Matsuura B, Dong M, Miller LJ, Matsui H, Onji M: **Molecular characterization and distribution of motilin family receptors in the human gastrointestinal tract.** *J Gastroenterol* 2006, **41**:223-230.
59. Bassil AK, Dass NB, Murray CD, Muir A, Sanger GJ: **Prokineticin-2, motilin, ghrelin and metoclopramide: prokinetic utility in mouse stomach and colon.** *Eur J Pharmacol* 2005, **524**:138-144.
60. Tan CP, Sano H, Iwaasa H, Pan J, Sailer AW, Hreniuk DL, Feighner SD, Palyha OC, Pong SS, Figueroa DJ, Austin CP, Jiang MM, Yu H, Ito J, Ito M, Guan XM, MacNeil DJ, Kanatani A, Van der Ploeg LH, Howard AD: **Melanin-concentrating hormone receptor subtypes 1 and 2: species-specific gene expression.** *Genomics* 2002, **79**:785-792.
61. Joost P, Methner A: **Phylogenetic analysis of 277 human G-protein-coupled receptors as a tool for the prediction of orphan receptor ligands.** *Genome Biol* 2002, **3**:RESEARCH0063.
62. **The NCBI BLAST archive** [<http://www.ncbi.nlm.nih.gov/BLAST/download.shtml>]
63. **The NCBI databases** [<http://www.ncbi.nlm.nih.gov/Database/>]
64. Kent WJ: **BLAT--the BLAST-like alignment tool.** *Genome Res* 2002, **12**:656-664.
65. **Celera** [<http://www.celera.com>]
66. **Entrez** [<http://www.ncbi.nlm.nih.gov/Entrez>]
67. Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite.** *Trends Genet* 2000, **16**:276-277.
68. **The HUGO Gene Nomenclature Committee** [<http://www.genenames.org>]
69. Rodriguez I, Del Punta K, Rothman A, Ishii T, Mombaerts P: **Multiple new and isolated families within the mouse superfamily of V1r vomeronasal receptors.** *Nat Neurosci* 2002, **5**:134-140.
70. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M: **Crystal structure of rhodopsin: A G protein-coupled receptor.** *Science* 2000, **289**:739-745.
71. Marchler-Bauer A, Bryant SH: **CD-Search: protein domain annotations on the fly.** *Nucleic Acids Res* 2004, **32**:W327-31.
72. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673-4680.
73. Felsenstein J: **Phylogenetic inference package.** 3.6th edition. Washington, Department of Genetics, University of Washington; 1993.
74. Schmidt HA, Strimmer K, Vingron M, von Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.** *Bioinformatics* 2002, **18**:502-504.
75. Foord SM, Bonner TI, Neubig RR, Rosser EM, Pin JP, Davenport AP, Spedding M, Harmar AJ: **International Union of Pharmacology. XLVI. G protein-coupled receptor list.** *Pharmacol Rev* 2005, **57**:279-288.
76. Lee CW, Rivera R, Gardell S, Dubin AE, Chun J: **GPR92 as a new G12/13- and Gq-coupled lysophosphatidic acid receptor that increases cAMP, LPA5.** *J Biol Chem* 2006, **281**:23589-23597.
77. Wang J, Simonavicius N, Wu X, Swaminath G, Reagan J, Tian H, Ling L: **Kynurenic acid as a ligand for orphan G protein-coupled receptor GPR35.** *J Biol Chem* 2006, **281**:22021-22028.
78. Wang J, Wu X, Simonavicius N, Tian H, Ling L: **Medium-chain fatty acids as ligands for orphan G protein-coupled receptor GPR84.** *J Biol Chem* 2006, **281**:34457-34464.
79. Kohno M, Hasegawa H, Inoue A, Muraoka M, Miyazaki T, Oka K, Yasukawa M: **Identification of N-arachidonylglycine as the endogenous ligand for orphan G-protein-coupled receptor GPR18.** *Biochem Biophys Res Commun* 2006, **347**:827-832.
80. Ignatov A, Robert J, Gregory-Evans C, Schaller HC: **RANTES stimulates Ca2+ mobilization and inositol trisphosphate (IP3) formation in cells transfected with G protein-coupled receptor 75.** *Br J Pharmacol* 2006, **149**:490-497.
81. Rezgaoui M, Susens U, Ignatov A, Gelderblom M, Glassmeier G, Franke I, Urny J, Imai Y, Takahashi R, Schaller HC: **The neuropeptide head activator is a high-affinity ligand for the orphan G-protein-coupled receptor GPR37.** *J Cell Sci* 2006, **119**:542-549.
82. Kotarsky K, Boketoft A, Bristulf J, Nilsson NE, Norberg A, Hansson S, Owman C, Sillard R, Leeb-Lundberg LM, Olde B: **Lysophosphatidic acid binds to and activates GPR92, a G protein-coupled receptor highly expressed in gastrointestinal lymphocytes.** *J Pharmacol Exp Ther* 2006, **318**:619-628.
83. Overton HA, Babbs AJ, Doel SM, Fyfe MC, Gardner LS, Griffin G, Jackson HC, Procter MJ, Rasamison CM, Tang-Christensen M, Widdowson PS, Williams GM, Reynet C: **Deorphanization of a G protein-coupled receptor for oleoylethanolamide and its use in the discovery of small-molecule hypophagic agents.** *Cell Metab* 2006, **3**:167-175.
84. Filardo E, Quinn J, Pang Y, Graeber C, Shaw S, Dong J, Thomas P: **Activation of the novel estrogen receptor, GPR30, at the plasma membrane.** *Endocrinology* 2007.
85. Sugo T, Tachimoto H, Chikatsu T, Murakami Y, Kikukawa Y, Sato S, Kikuchi K, Nagi T, Harada M, Ogi K, Ebisawa M, Mori M: **Identification of a lysophosphatidylserine receptor on mast cells.** *Biochem Biophys Res Commun* 2006, **341**:1078-1087.
86. **TMHMM** [<http://www.cbs.dtu.dk/services/TMHMM>]