Evolutionary Bioinformatics

ORIGINAL RESEARCH

Cross-Genome Clustering of Human and *C. elegans* G-Protein Coupled Receptors

Balasubramanian Nagarathnam[1,2], Singaravelu Kalaimathy[1], Veluchamy Balakrishnan[2] and Ramanathan Sowdhamini[1]

[1]National Centre for Biological Sciences, Tata Institute of Fundamental Research, GKVK Campus, Bellary Road, Bangalore 560065, India. [2]Department of Biotechnology, K.S.Rangasamy College of Technology, K.S.R. Kalvi Nagar, Tiruchengode 637215, India. *Corresponding author email: mini@ncbs.res.in

Abstract: G-protein coupled receptors (GPCRs) are one of the largest groups of membrane proteins and are popular drug targets. The work reported here attempts to perform cross-genome phylogeny on GPCRs from two widely different taxa, human versus *C. elegans* genomes and to address the issues on evolutionary plasticity, to identify functionally related genes, orthologous relationship, and ligand binding properties through effective bioinformatic approaches. Through RPS blast around 1106 nematode GPCRs were given chance to associate with previously established 8 types of human GPCR profiles at varying *E*-value thresholds and resulted 32 clusters were illustrating co-clustering and class-specific retainsionship. In the significant thresholds, 81% of the *C. elegans* GPCRs were associated with 32 clusters and 27 *C. elegans* GPCRs (2%) inferred for orthology. 177 hypothetical proteins were observed in cluster association and could be reliably associated with one of 32 clusters. Several nematode-specific GPCR clades were observed suggesting lineage-specific functional recruitment in response to environment.

Keywords: G-protein-coupled receptors (GPCRs), serpentine receptors, PSSM profile, cross-genome clustering, phylogeny

Evolutionary Bioinformatics 2012:8 229–259

doi: 10.4137/EBO.S9405

This article is available from http://www.la-press.com.

Introduction

G-protein-coupled receptors (GPCRs) belong to a functionally diverse superfamily engaged in cellular activities, in particular to signal transmission.[1] These membrane proteins are ubiquitous[2] and proved to be very successful drug targets due to diverse ligands of varying sizes.[3] Substantial evidence on GPCR oligomerization,[4] participation in signaling pathways,[5] clinical importance[6] and availability of repositories for multiple organisms[7] provide significant reasons to perform sequence analysis on cross-genome GPCRs. The significant genetic tractability in the whole genome of *D. melanogaster* (1%), *C. elegans* (<5%) and occurrence of more than 1000 candidate GPCRs in humans[1,7,8] are further reasons to investigate these cell surface receptors across genome. Our previous publication proposed a novel approach to establish phylogenetic cluster association to eight major groups of human and *D. melanogaster* GPCRs,[9] whereas in this paper we are interested in dealing with another animal model *C. elegans*[10] to associate it with already grouped known human GPCR cluster dataset by generating human GPCR profiles (PSSM) for the representative sequence from the known human GPCR cluster dataset and performing profile based clustering technique (RPS-BLAST)[11–13] for unknown association.

In principle, the cross-genome phylogenetic approach is aimed at understanding the evolutionary plasticity to discover functional similarities and to identify functionally related genes,[14] orthologous relationships[15] and preserved motif patterns across these two organisms. The current study, in turn, will enable us to observe the details of cluster association in retaining nematode-specific gene clusters and also the established evolutionary integrity with human GPCRs at cross-genome phylogeny.

The selected model organism *C. elegans* carries an extensive repertoire of Pfam domain matches, conserved signalling pathways and homologues of proteins found in other organisms, such as human and *Drosophila*.[6] Further, there is an increasing evidence for genetic and physiological similarity (*eg,* stress response and basic physiological processes) with higher order organisms (humans) and conservation of 12 out of 17 known signal transduction pathways (like SynMuv Pathway) in *C. elegans* and humans[16,17] are note worthy to deal with *C. elegans* GPCRs to explore the nematode genome for its genetic influence in the evolutionary trends[18] and the effectiveness of employing *C. elegans* as a model organism for understanding fundamental pathways in higher order organisms. Already, evidence is emerging that genetic defects observed in *C. elegans* have counterparts in human diseases and *C. elegans* might provide information about cellular processes affecting human diseases.[6]

C. elegans chemoreceptor genes are strikingly abundant and diverse, possessing around 400 apparent pseudogenes and almost ~1300 predicted genes that encode members of putative chemosensory genes[19,20] and usually referred to as serpentine receptors (SR) or chemosensory receptors (CR). These genes are classified under serpentine receptors (SR) superfamily[21–23] and about 7% of occurrence of serpentine receptors in the whole genome indicates the extreme dependency of chemosensory abilities due to the the absence of visual and auditory systems in *C. elegans.*

Through genetic screening, Sengupta *et al* identified odr-10 in *C. elegans* participating in olfactory response which in turn reveals the relationship between odr-10 and other serpentine receptors in *C. elegans* GPCR families.[24,23] Recent reports are stating that serpentine receptors like srg 36, srg 37 are pheromone like receptors and are participating in sensing ascaroside pheromones which are observed in sex chromosomes.[25] In other instances that srbc-64 and srbc-66 candidate serpentine receptors are responsible for pheromone activity in *C. elegans*.[19,26] These case studies are helpful in understanding the divergent chemoperception properties in *C. elegans* with reference to GPCR.

A detailed compilation on SR superfamilies and relative families of odr-10 by Robertson and co-workers.[23,19,27] provide information on phylogenetic distribution, function and expression patterns in *C. elegans* chemosensory receptors. They classify *C. elegans* serpentine receptors into nearly 20 recognizable families (refer Table S1) on the basis of sequence similarity and shared intron locations. 19 of these families are well-established and grouped under superfamilies such as Sra superfamily (sra, srab, srb, and sre), Srg superfamily (srg, srt, sru, srv, srx, and srxa), Str super family (srd, srh, sri, srj, str) and others or Solo type includes srbc, srsx, srw and srz. Notably the large Str family along with related sri and srj families are observed to be related to odr-10 (olfactory receptor) in *C. elegans*.[23,19,27] The occurrence of gene duplication, redundancy, movement and diversification of genes in *C. elegans*[28] insist on the need for careful assignment of each gene description, particularly while referring to phylogenetic clustering. This abundant genetic investment in *C. elegans* was utilized in our study for the cross—genome clustering with selected human GPCRs,[9] and we were interested to observe the inter-genomic clustering of these SR types, when we introduce previously annotated and associated human GPCRs, in turn, a strong evolutionary pressure.

Materials and Methods

Data

Collection of *C. elegans* GPCR sequences

The draft GPCR sequences for the complete genome of *C. elegans* was obtained from SEVENS database[29] and around 1204 sequences were collected and predicted for the number of transmembrane helices (TM helices) in each sequence for membrane topology (refer step 1.1 in Fig. 1).

Prediction of membrane topology for *C. elegans* GPCRs

The membrane topology of each GPCR sequence was predicted by using SOSUI[30] and HMMTOP[31] prediction methods (refer step 1.2 in Fig. 1). The observed consensus from both methods was used to define the eligible candidate GPCRs (refer Table S2).

Removal of *C. elegans* GPCRs for over/under predicted TM helices

The *C. elegans* GPCR sequences predicted for 7 (±2) TM helices were retained whereas GPCRs predicted to lower or upper to the mentioned cut-off was removed from the dataset. Thus totally 1160 GPCR sequences of *C. elegans* were retained after this screening procedure[30,31] and retained as "*C. elegans* GPCR dataset" (refer step 1.3 in Fig. 1).

Human and *Drosophila* GPCR cluster dataset (pre-aligned set of GPCR cluster association)

Human-*Drosophila* GPCR cluster dataset for 32 clusters of eight major groups were obtained from our previous publication[9] (herein we refer as "known cluster association"). Since we were interested to perform cross-genome phylogeny with *C. elegans* GPCRs this time, the associated *Drosophila* GPCR sequences were eliminated from the previously established human-*Drosophila* GPCR cluster so to get human GPCR-only clusters from 31 clusters and *Drosophila* GPCR-only cluster from one cluster (cluster no: 26, since it was associated only with *Drosophila* GPCRs[9]) and has been used for our current study.

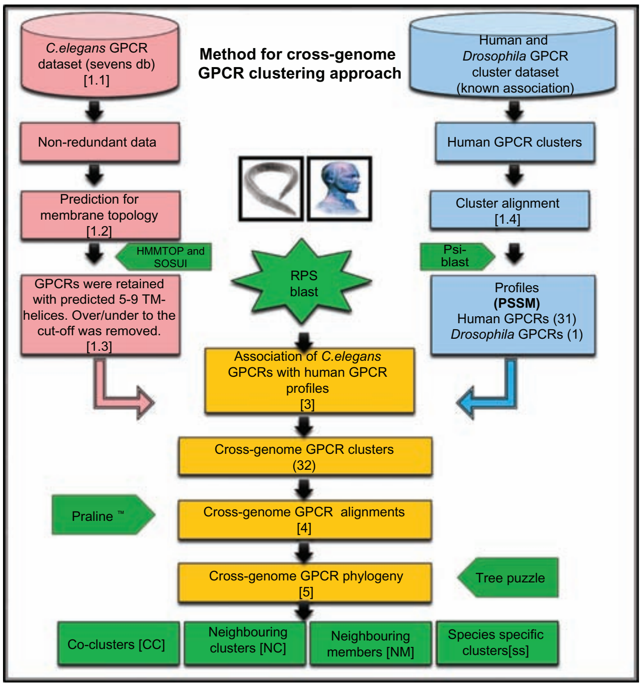
C.elegans GPCR dataset (sevens db) [1.1]
Method for cross-genome GPCR clustering approach
Human and Drosophila GPCR cluster dataset (known association)
Non-redundant data
Human GPCR clusters
Prediction for membrane topology [1.2]
Cluster alignment [1.4]
HMMTOP and SOSUI
RPS blast
Psi-blast
GPCRs were retained with predicted 5-9 TM-helices. Over/under to the cut-off was removed. [1.3]
Profiles (PSSM) Human GPCRs (31) Drosophila GPCRs (1)
Association of C.elegans GPCRs with human GPCR profiles [3]
Cross-genome GPCR clusters (32)
Praline ™
Cross-genome GPCR alignments [4]
Cross-genome GPCR phylogeny [5]
Tree puzzle
Co-clusters [CC]
Neighbouring clusters [NC]
Neighbouring members [NM]
Species specific clusters[ss]

Figure 1. Flowchart describing steps involved in the study for cross-genome clustering of human and *C. elegans* G-protein coupled receptors. Detailed procedure on each step was explained in Methods.

Due to the removal of *Drosophila* GPCR sequences from the human-*Drosophila* GPCR cluster dataset of the known cluster association (except for 26th cluster), the extra indels were observed in the previous alignments. The observed extra indels in each alignment position of the MSA were carefully edited manually by using MEGA 4.0.[36] The resulting improved alignment for 32 clusters were retained with totally 353 human GPCR sequences from 31 clusters and 14 *Drosophila* GPCR from one cluster (in Cluster 26) to get 32 clusters. (refer step 1.4 in Fig. 1). We refer these pre-aligned set of GPCR association as "GPCR cluster dataset".

Generation of representative profiles

The previously aligned GPCR cluster dataset (by ClustalW) for 32 clusters of known receptor types were used to generate position-specific scoring matrix

(PSSM) or profile[11–13] to represent cluster/receptor specific sequence properties. In our current attempt, for 32 clusters, 32 Profiles were created by supplying their respective MSA of a representative sequence to Psi-Blast procedure (Position-Specific Iterative Basic Local Alignment Search Tool)[32] ultimately to produce PSSM-profiles of 32 Cluster specific profiles eight major receptor types. We refer these profiles as "representative profiles", since it represents sequence property of respective alignments from 32 clusters (refer step 2 in Fig. 1).

Performing RPS-BLAST

Trial study with known associations

Separately, *E*-value thresholds were standardized by performing a trial study with randomly collected 102 *Drosophila* GPCR sequences from the known cluster association[9] and *Drosophila* GPCRs are given chance to pick appropriate human GPCR profile [refer step 2] as a first hit by using RPS-BLAST. By running RPS-BLAST, we could observe around 102 *Drosophila* GPCR queries picking up correct cluster association in terms of respective human GPCR profile, correct cluster number and receptor type at the significant *E*-value cut-off <0.001 (refer Table 3) and 11 *Drosophila* GPCR queries were observed for picking up for the same receptor type, but selecting different cluster numbers. This could be explained by observing that the sequence property of receptor types at the cluster level is interconnected, since the sequences related to the same receptors were distributed into various sub-clusters (for example Peptide type receptors are distributed in to 11 clusters and Chemokine type receptors are of 2 clusters and so on in the dataset[9]). This could be the reason that the 11 sequences are observed to associate with their same receptor type but different cluster number (for example: the Q8MKUO receptor picks up Cluster 8, a peptide receptor, but the expected association also from a peptide type receptor, but from Cluster 11.

The failure to pick up respective representative profile may be due to cross-cluster association at Known GPCR association from the dataset.[9] Also, relaxed *E*-value thresholds >0.001 can be another reason for providing the freedom of selecting nearest PSSM profiles. This pilot study clearly shows diversity of sequence properties for the same receptor type across clusters [refer Table S3] within the known human GPCR dataset. With this significant initial standardization and confidence level of 90 % for correct association the queries from *C. elegans* GPCR dataset (unknown association) were given chance to select its mostly related GPCR profile from the generated human (31)/*Drosophila* (1) GPCR cluster dataset by employing RPS-Blast (Reverse PSI-Blast).

Setting *E*-value thresholds for unknown association

In preliminary analysis, correct associations between the 100 selected *Drosophila* GPCRs and their respective human GPCR profiles were observed at *E*-value range of <0.001 (refer Table S3 for 100 *Drosophila* GPCRs). Various ranges of *E*-value thresholds from 0.001, 0.01, 0.5, 1.0 and <5.0 to >5.0 till 14 were tried with *C. elegans* GPCRs to select its closely related profile from the dataset with little chance of encountering false connections (refer Table S3 for the 11% of predicted false association for Known Cluster association). A need to relax the *E*-value thresholds for different ranges has arisen due to fact of higher evolutionary divergence between human and *C. elegans* for the current study.

Choice of picking the closest representative profile by each *C. elegans* GPCR sequence at varying *E*-values was guiding to assign the respective *C. elegans* GPCRs to the respective human GPCR cluster which has previously associated sequences. The cross genome cluster association was decided on the basis of its respective profile, and significant bit score, percentage identity and *E*-value from the hit list arrived from RPS-BLAST. The *E*-value thresholds are mainly considered for finalizing the association and the details are given in Table S4. Then to organize cross-genome GPCR cluster alignments, the previously associated human GPCR sequences of the respective known human GPCR profile was aligned with newly associated *C. elegans* GPCR sequences (refer step 3 in Fig. 1).

Cross–genome alignment of human—*C. elegans* GPCRs

The pre-aligned set of GPCR sequences from the dataset with newly associated *C. elegans* GPCR sequences were aligned by an appropriate multiple alignment tool called PRALINE[TM,33] to generate cross-genome sequence alignments of 32 GPCR

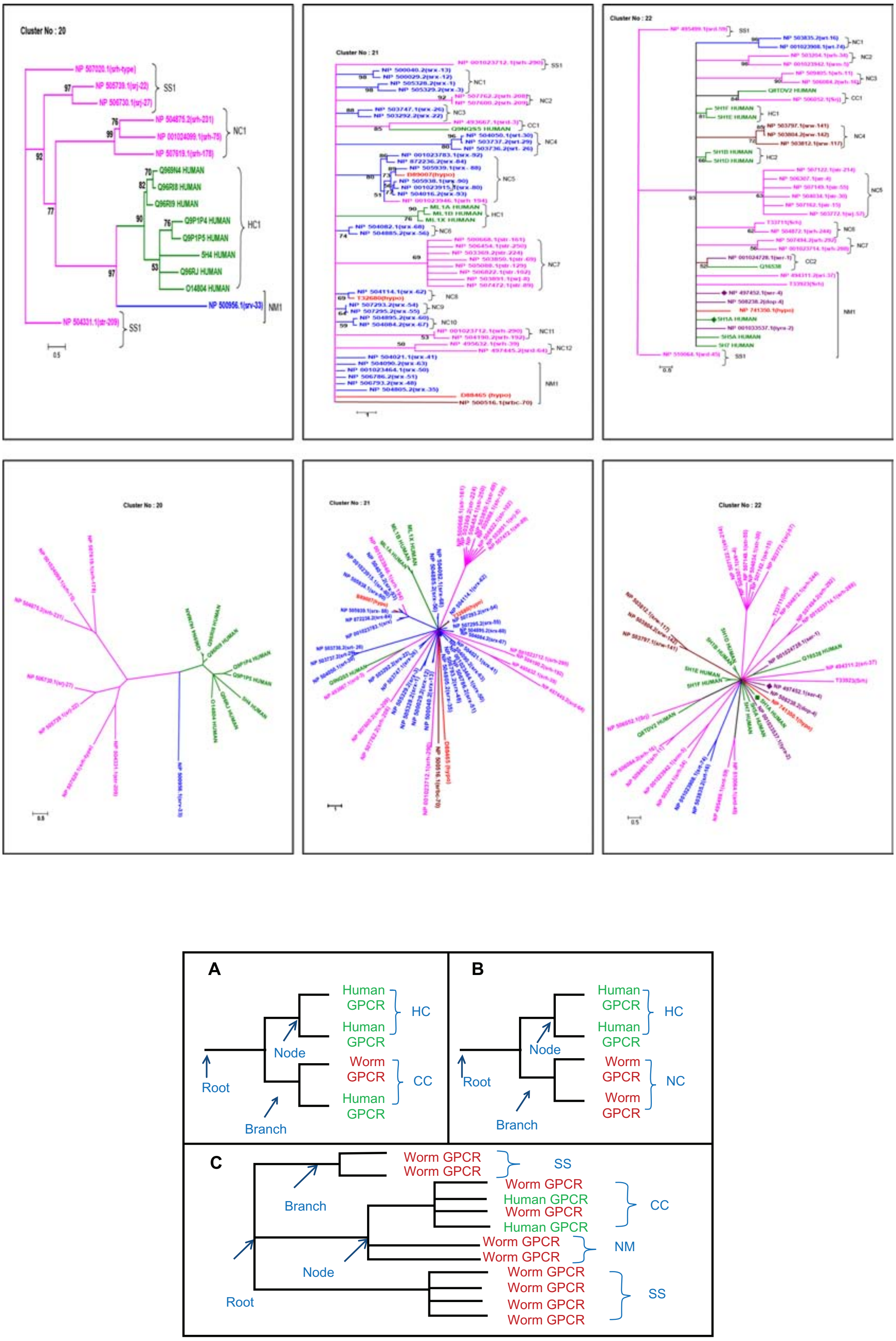

Figure 2. Illustrative phylogenetic tree depicts the clustering of biogenic amine receptors (Clusters 20–22) in rectangular (Top) and radial (Bottom) displays also a pictorial representation of tree topologies has been given to guide the types of cluster association. (**A**) denotes the cluster association for HC and CC, wherein HC refers to the association of GPCRs from only human genome, CC refers to the inter genomic cluster association of GPCRs from both human and *C.elegans*. (**B**) refers to the NC association, where the occurrence of *C.elegans* GPCRs has observed at adjacent or neighbouring to HC. (**C**) refers to the association for species specific (SS), Co-clusters (CC) and neighbor member (NM) occur in the tree topology. (Refer section "Terminologies used to describe phylogeny" also).

clusters. In principle, the alignment is based on the TM topology and profile scoring schemes such as PHAT matrix with the gap penalty of 15 (1 extension) for predicted TM regions and Blosum 62 matrix for non-TM regions with the gap penalty of 16.5 (1 extension) followed by an iterative scheme to enhance the alignment quality. Where necessary, alignments were further optimized by manual editing through MEGA 4.0 software[36] (refer step 4 in Fig. 1).

Cross genome phylogeny of human—*C. elegans* GPCRs

The generated cross-genome GPCR cluster alignments of human—*C. elegans* GPCRs were clustered to generate cross-genome phylogeny by using Tree-Puzzle.[34,35] This package was employed to perform quartet-based maximum-likelihood phylogenetic analysis with the puzzling step of 10,000 times and resolved trees were generated as a phylogenetic tree file (outtree files). The resultant tree files were viewed by MEGA 4.0.[36] Some critical decisions of the constructed phylogenetic trees were done based on branching patterns and cluster association (refer Fig. 2A–C). Also, the bootstrap supporting values (herein referred as bs), cluster association of serpentine receptor type (superfamily level) are observed both a traditional rectangular (known as dendrogram) and radial view (known as radial display) for the analysis and final graphical display and result interpretations (refer step 5 in Fig. 1). Rectangular display enables better understanding of the distribution of *C. elegans* GPCRs for various types of cluster associations and their nomenclature. Radial display provides an overview of the distribution of *C. elegans* GPCRs with respect to receptor superfamilies and this will give immediate understanding on cluster distribution of distantly related members in cross-genome phylogeny.

Results

In the current study (as mentioned in Methods), 1106 *C. elegans* GPCR sequences were associated by querying against the database of human GPCR profiles from the known cluster association by using a sensitive RPS-Blast technique.[11–13] The newly associated *C. elegans* GPCR sequences were tabulated (Table 1) for clusterwise association with significant *E*-values (Table S4) using the procedure for cross-genome association discussed detailed in Methods (also refer Fig. 1).

Table 1. Distribution of human and *C. elegans* GPCRs in 32 Clusters.

Cluster no.	**Receptor type**	**No. of human GPCRs**	**No. of *C. elegans GPCRs***
1	PR	8	32
2	PR	11	40
3	PR	8	34
4	PR	8	32
5	PR	8	54
6	PR	8	42
7	PR	8	34
8	PR	8	34
9	PR	14	54
10	PR	8	26
11	PR	12	60
12	CMK	10	17
13	CMK	15	13
14	N&L	7	64
15	N&L	18	53
16	N&L	7	21
17	N&L	18	30
18	N&L	8	28
19	N&L	18	89
20	BGA	8	8
21	BGA	4	49
22	BGA	8	30
23	BGA	22	77
24	BGA	23	68
25	SEC	16	29
26	SEC	0	20
27	CAR	29	16
28	GLR	8	20
29	GLR	5	14
30	GLR	4	23
31	GLR	3	8
32	FRZ/SMT	11	40

Note: List of cluster wise distribution (for 32 clusters) of *C. elegans* GPCRs according to the eight subtypes of human GPCRs.

The association observed between the two genomes is considered for the best connecting sequences for known receptor type in higher–order organism with a effective model organism further to compare the sequence properties, then to connect the structural, functional, and evolutionary relatedness among them.

Terminologies used to describe phylogeny

The resulting 32 different tree topologies of cross-genome phylogeny describes five possible types of association and are discussed below (refer to Fig.2. A–C for the parts of cross-genome phylogenetic tree to denote inter and intra genomic level of association in the tree topology).

Human GPCR clade [HC]: refers to the pure distribution (homogenous occurrence) of human GPCRs in the established branching pattern at intra genomic level and are referred as HC in tree topology (refer Fig. 2A).

Co-clusters [CC]: refers to the co-clustering or heterogenous distribution or clear intermixing of *C. elegans* GPCR*s* with human GPCRs to represent a strong cross-genomic clustering in the established branching pattern/clade at intergenomic level and is denoted as ‘CC’ in phylogenetic tree (refer Fig. 2A).

Neighbour Clades [NC]: refers to homogenous occurrence of *C. elegans* GPCR*s* adjacent or neighbouring human GPCRs clusters [HC] in the established branching patterns at intra genomic level and is referred as ‘NC’ in phylogenetic tree (refer Fig. 2B).

Neighbour Members [NM]: refers to pure (homogenous) or intermixed (heterogenous) distribution of GPCRs in the branching pattern with limited nodes, mostly originating from root and is denoted as ‘NM’ (refer Fig. 2C). However, the observed associations may not be viewed as closely related at intergenomic level as mentioned in CC.

Species specific members [SS]: refers to pure distribution (homogenous) of *C. elegans* GPCRs and remains as separate clade in a tree topology. The clades are denoted as ‘SS’ in phylogenetic tree. In this current study, although we refer species-specific members, it is only in the context of cross genome human—*C. elegans* GPCRs and this term does not imply complete set of unique genes observed only in one species in the entire taxonomy and evolutionary tree of life (refer Fig. 2C).

Superfamilies of serpentine receptors (SR): The distribution of chemosensory receptors of *C. elegans* are discussed according to superfamilies like Sra, Str, Srg and Other, broken into 24 types, as suggested by Robertson and coworkers.[23,19,27,37] This classification has been followed throughout the 32 clusters to appreciate the influence of human GPCRs and species-specific preservation of such superfamily members.

Cluster-wise summary: The observed cross-genome cluster associations were discussed in detail below with a cluster-wise summary according to the observed cross-genome GPCR topology/phylogeny.

Peptide receptors (PR)

Clusters 1–11 are related to peptide receptors and around 442 GPCRs from *C. elegans* could be associated with nearly 101 peptide receptors of humans in the dataset (refer Table 1). PR occurs predominantly in the dataset. Generally, the size of the peptide ligands varies from two amino acid residues to as many as 50. Broadly GPCRs occur in A1–A19 sub groups. Many of the peptide receptors are related to potential clinical applications and related to various diseases such as chronic inflammatory diseases, degenerative diseases, autoimmune diseases, cancer, cardiovascular diseases etc,[9] hence these receptors also act as interesting drug targets.

Few important receptors are given along with their group to emphasize the distribution of various PR in the clusters 1–11. Small peptide receptors such as angiotensin (8 amino acids), bradykinin (9 amino acids)—(Cluster 3), apelin (`~36 amino acids) and orphan receptors GPRF that act as co-receptor for the human immunodeficiency virus (HIV),[45] belong to group A3 as observed in the cluster dataset. The receptors related to group A4 such as opioid and somatostatin receptors are found in Cluster 2.

Galanin receptor, cysteinyl leukotriene receptor, leukotriene B4 receptor, relaxin receptor, melanin-concentrating hormone receptor, KiSS1-derived peptide receptor (GPR54),[38] urotensin-II receptor (refer Cluster 1) related to group A5 are also observed in the dataset.

Cholecystokinin receptor, neuropeptide FF receptor, orexin receptor, vasopressin receptor, gonadotrophin releasing hormone receptor (GNRHR, GRHR) belongs to group A6 and are seen in Cluster 6.

Neuromedin U receptor (Cluster 5), neurotensin receptor (Cluster 5), thyrotropin-releasing hormone receptor (TRHR, TRFR), GHSR, GPR39, GHSR, GPR39 (Cluster 5), bombesin receptor, gastrin-releasing peptide receptor (GRPR), endothelin receptor (Cluster 7), thyrotropin-releasing hormone receptor (TRHR, TRFR), and motilin receptor that are related to group A7 and are observed in the cluster dataset.

Anaphylatoxin receptors, formyl peptide receptor, MAS1 oncogene, GPR1 (Cluster 6), GPR32 (Cluster 9), GPR44 and GPR77 (Cluster 9) belong to Group A8, while neurokinin receptor (Cluster 11), neuropeptide Y receptor (Cluster 11), prolactin-releasing peptide receptor, prokineticin receptor 1,2 (Cluster 11), GPR19, GPR50, GPR75 and GPR83 are from group A9. Glycoprotein hormone receptor, leucine-rich repeat-containing G protein-coupled receptor 4, LGR5, LGR6 from group A10 (Cluster 7) are present in the cluster associations, further allowing to explore the functional relevance with nematode GPCRs.

Cluster 1

Cluster 1 are associated with eight human peptide receptors and 32 *C. elegans* GPCRs at different *E*-value cutoffs (refer Table S4). Human peptide receptors are distributed into clades, having only human GPCRs (HC1) and another clade having human GPCRs with a *C. elegans* GPCR (CC1). Interestingly in CC1, co-clustering of neuropeptide receptor (npr-9) from *C. elegans* with human GPCR (Q969F8) was observed (Fig. S1). The human GPCR (Q969F8), observed in CC1, is implicated in breast carcinomas and is also majorly involved in endocrine regulations and onset of puberty.[37] Mutations on this gene have been associated with hypogonadotropic hypogonadism and central precocious puberty in humans.[38–40] Such functionally important human peptide receptor has an orthologous relationship (refer Table S5) with neuropeptide receptor (npr-9) of *C. elegans* and proves to be an interesting target to be studied in *C. elegans* model organism. Neighbouring clusters (NC1–NC5) exhibit topology only with *C. elegans* GPCRs, where NC1–NC3, NC5 also form neighbor members and NM2 includes a majority of hypothetical proteins. This further helps to interpret that these unknown or unannotated GPCRs (refer Table S5) in nematode are probably related to human peptide receptors through cross-genome phylogeny also probably belong to class A GPCRs (Rhodopsin-like). NC4 associates with pure set of srd members. Receptors from Str superfamily (str, srh type) are observed as neighbour members in NM1.

SS1 and SS2 retain candidate receptors from srw type[19] in *C. elegans*. srw type is related to families of FMRF-amide and other peptide receptors, which are expected to have relatives in vertebrates and insects and have strong clustering at the chromosomal level.[23] Their associations with Cluster-1 human peptide receptors suggest that they serve as closely related environmental peptide receptors.

Cluster 2

11 candidate human peptide receptors (HC1) and 41 serpentine receptors of *C. elegans* are distributed in Cluster-2 (Fig. S1). HC1 remains as clade with a majority of somatostatin receptor types[41] and the neighbouring clade NC1 has members from Str superfamily (sri, str). NC2, NC4 and NC5 retain receptors of srsx, srv, srsx and srw family,[23] suggesting clear superfamily conservation. NC6 branched with srm, srx ('Others' superfamily) with a hypothetical receptor and NC7 includes six hypothetical proteins[42] including an unidentified vitellogenin-linked transcript family member (uvt-6). Particularly, uvt-6 branches with a hypothetical protein (NP_510833.3) that belongs to rhodopsin family and is most similar to the mammalian somatostatin receptors (referWBGene00006864).[43] This helps us to understand the association of *C. elegans* GPCR members with human peptide receptors by RPS-BLAST where the somatostatin receptor types are present in its profile; this association is also observed in SEVENS database.[29] Diverse members, like of srd, srv, sre types, are also observed as neighbouring members in NM1. The observed pure dispersion of srh type receptors of *C. elegans* at SS1 indicates the high nematode–specific tendencies for these receptor types.

Cluster 3

In Cluster-3, HC1 retains eight entries of human peptide receptors. 34 candidate GPCRs from *C. elegans*[44] are distributed into neighbouring clusters NC1–NC5, neighbouring members NM1 with 16 mixed receptor types and species-specific clade SS1 (Fig. S1). HC1 remains with closely related peptide receptors like Angiotensin II receptor, Bradykinin receptor I, GPR15 types belonging to the same subfamily of A3 in GPCR Classification.[45]

NC1 retains candidates from srd type of Str superfamily. Srd types of chemoreceptors in *C. elegans* retain one potential disulfide bond in extracellular domain 2 and shares with the srh and sri families a highly conserved PYR sequence at or near the inner end of transmembrane (TM) helices 7 and belongs to Pfam profile PF10317.[46] NC2 carries srxa and str type *C. elegans* GPCR members and NC3–NC5 clades carry pure set of serpentine receptors of sri, hypothetical protein and srw members, respectively. Abundant srw members with comparable representation from Str superfamily (srh type) and Srg superfamily (srt, srv type) of *C. elegans*[23] are dispersed as neighbouring members in NM1.SS1 covers six candidate GPCRs from Str superfamily and single representation from Srg superfamily.

Cluster 4

In Cluster-4, 32 candidate GPCRs from *C. elegans* and 8 human GPCRs were observed and subgrouping of human GPCRs into HC1–HC3 helps to understand evolutionary integrity at intragenomic level (Fig S2). HC1 includes human Gastrin-releasing peptide receptor (GRPR), neuromedin-B receptor (NMB-R) (Neuromedin-B-preferring bombesin receptor), related to Lung carcinoma,[47] and BRS3 (Bombesin receptor subtype-3) related to obesity and associated to diseases,[48] HC2 retains functionally related Endothelin receptor, Non-selective type (ETBR) and Endothelin-1 receptor precursor; ET-A (ET1R).

G-protein-coupled receptor 37 and endothelin receptor type B-like[49] are noticed in HC3 carrying the same functional property, such as participating in glucocorticoid actions and blood pressure control. CC1 refers the co-clustering of human GPCR (Q8TDVO) and a *C. elegans* receptor (NP_502893.2 (srv-14), refer: WBGene00005725) which illustrates the intergenomic association at cross-genome phylogeny (Fig. S2), suggesting similar function. The neighbouring clusters include srsx, srw and srh types and are distributed in NC1, NC3 and NC4, respectively in a common fashion covering two members of the same receptor type. NC2 retains sri type receptor with a hypothetical protein. NC5 clade includes two members of srj type of Str superfamily with a hypothetical protein. NC6 includes members of hypothetical proteins (Rhodopsin-like), NC7, NC8, NC11 retain srh type members uniquely. NC10 includes candidates from srh type and an unannotated transmembrane protein. And two srh type receptor, with a hypothetical protein observed in NM1 further to connect the functional relevance. Overall, the cluster is an illustrative model in explaining the rich distribution of Str type of SR with hypothetical proteins to connect functional relevance from known receptor type. SS1 includes four srd members of Str superfamily.

Cluster 5

Cluster-5 carries 54 *C. elegans* GPCRs and 8 human GPCRs (Fig. S2) and dispersion of human GPCRs in CC1 is notable for cross-genome clustering, wherein human thyrotropin-releasing hormone receptor (TRFR_HUMAN) coclusters with a GPCR from *C. elegans* (NP_491990.1-rhodopsin-like/hypothetical protein). Indeed, such co-clustering is expected in this case, since there is significant evolutionary similarity between the two proteins and the pair is termed as an "ortholog" (refer Table S5) also shares similar functional cues in calcium signaling pathway [KEGG PATH: Ko04020].

Further, NP_505077.1 (str-138) Neuromedin U receptor 2 of *C. elegans* is observed to retain orthologous relationship with human GPCR-Q96 AM5 although their functional equivalence is yet to be established.

NC1 to NC10 are reported as neighbouring clusters, wherein NC1 includes candidates of Sri, Srh type from Str superfamily and uniformly the pure set of Srx and Srw candidates are observed in NC2 and NC6, whereas rest of the neighbouring clusters (NC3–NC5, NC7–NC10) retain mostly hypothetical proteins/receptors. NC8 clade branches with GNAT (GCN5-related N-acetyltransuperfamilyerase (GNAT) family protein) at the bs of 59 with hypothetical protein member and NC9 is branching with a Spr-2 "Sex Peptide Receptor (*Drosophila*) Related family member (sprr-2)" (NP_510455.2). Notably, all the neighbour members in NM1 are also from hypothetical proteins. Overall, this cluster covers majorly of Rho-like members (hypothetical proteins) and provides broader scope in connecting functional relevance with available 19 subgroups (A1–A19) of GPCRs of higher organisms.[50]

SS1 and SS2 include purely Str superfamily superfamily members (srj, str and srd) and notably SS3 retains species-specific olfactory receptor Odr-10[51] and branches with three other str type receptors. This superfamily helps to correlate the functional clues within candidate representation for olfaction in *C. elegans* and also proves the fact that Str/Stl family carries the large group of genes with special functional relationship towards Odr-10.[19] Since knowing Odr-10 is meant for olfactory perspective, association of this sequence with class-A member by RPS-BLAST, *albeit* at poor *E*-values and appearing as species-specific clade is encouraging suggesting that such cross-family connections can be recognized using significance of *E*-values and mode of clustering.

Cluster 6

Cluster-6 establishes a peculiar pattern of associating with eight hormone receptors of human with 42 *C. elegans* GPCRs. Limited branching patterns is the critical feature in describing this phylogeny suggesting polyploidy in this cluster (Fig. S2).

HC1 carries closely related V2R, V1BR, V2BR-Vasopressin receptors[47,53] and five other human GPCRs and are dispersed along with neighbour members in NM1 along with the diverse type of *C. elegans* GPCRs. NC1–NC5 are denoted as neighbouring clusters due to the occurrence of HC1 in between. NC1–NC3, NC5 clades, on the other hand retain only srw, srh, srsx, srh type members respectively. NC4 associates with a hypothetical protein and a gnrr (NP_491453.1) and around 27 entries are distributed as neighbour members in NM1 including five human peptide receptors. The neighbour members of *C. elegans* majorly includes candidates from srw type, five candidate GPCR from Str superfamily, six entries from Gnrr type, two entries of hypothetical proteins, and a sre type receptor. SS1 retains 2 srwreceptors in a distinct fashion to represent species-specific clade. This cluster retains a peculiar fashion of intermixing nematode gnrr (gonadotropin releasing hormone receptor)[52] with human gonadotropin releasing hormone receptor (GRHR) further helps to correlate biological significance in reproductive endocrinology in model organisms.

Cluster 7

In Cluster-7, eight human and 34 *C. elegans* GPCRs are associated, where the human GPCRs are branched into HC1 and HC2 suggesting distinct members observed even within the humanGPCRs clusters, whereas CC1 illustrates the co-clustering of leucine-rich repeat-containing GPCR7 (LGR7) and insulin-like peptide 3 receptors (LGR8)[54] of human GPCRs with a homologue from *C. elegans* (fshr-1) (Fig. S3). This strong co-clustering can be explained as being due not only due to the orthologous relationship to mammalian follicle stimulating hormone receptor, but also its functional importance in germline differentiation and survival in nematode taxon.[55] HC1 retains closely related human glycoprotein hormone receptors and HC2 branches with related human leucine-rich hormone receptors denoting the functional integrity observed at higher order organisms for these types of receptors.[9,45] This cluster covers species-specific members in SS1 including sri and srh type members of Str superfamily.[19] The neighbouring clusters NC2 and NC3 have candidate representation from srx, sre and srt type receptors. NC3–NC9 and following neighbouring members are abundant and purely distributed with members from the largest superfamily of Str and highly duplicated srt members of *C. elegans*.[19]

Cluster 8

Cluster-8 carries eight human GPCRs and 34 *C. elegans* GPCRs. HC1 retains neuropeptide receptors, HC2 also includes neuropeptide receptors related to wakefulness, food consumption, and locomotion in humans. Deletion of the orexin gene in mice produces a condition similar to canine and human narcolepsy in vivo.[56] CC1 includes human cholecystokinin (CCK) receptors, important for gall bladder contraction and pancreatic enzyme secretion[57] and is significantly branched with a nematode cholecystokinin receptor-type namely Ckr-2 (NP_001022842.1) at the bs of 55 (Fig. S3). The observed association further helps to analyse CCK receptors in two taxa for functional similarities and are illustrative of cross-genome association.

Eight neighbouring members of *C. elegans* GPCRs like srw, sprr, srxa type members from 'Other' type superfamily[19] two hypothetical protein members, a neuropeptide receptor (NK2R) and with other two human neuropeptide receptors suggesting a common functional relevance despite a heterogenous dispersion.

Neighbouring clusters are observed from NC1–NC3. NC1 is associated purely with the srd type receptors. NC2 includes a str candidate with a probable GPCR of *C. elegans* NP_509368.1 (C02B8.5). NC3 carries neuropeptide receptors (npr-1, npr-2) of *C. elegans*. Overall, Cluster 8 includes a huge number of srh type receptors from Str superfamily and the retention of this largest superfamily[19] is observed at nematode specific-clades from SS1–SS7 (Fig. S3) suggesting a significant over-representation or amplification of species-specific members.

Cluster 9

Cluster-9 has 54 entries of *C. elegans* GPCRs and 14 human GPCRs and the distribution of *C. elegans* Str superfamily receptors dominates along with srbc receptor population and equally intermixed human peptide receptors (Fig. S3). Human GPCRs within this cluster are sub-distributed into five clades (HC1–HC5) and also with *C. elegans* str type members. Four other human peptide receptors, observed as neighbour members as NM1 in tree topology, represent clear inter-genomic clustering. Interestingly, the formyl peptide receptors/chemoattractant receptors (FML1, FML2, FMLR and GP44) of human origin belonging to the subfamily of A8[50,58] remain as neighbour members, suggesting that *C. elegans* GPCR counterparts could be identified for such receptors. NC1–NC12 is observed for neighbouring clusters. Among 12 neighbouring clades (Fig. S3), particularly NC2, NC3, NC4, NC5, NC8, NC9, NC10 and NC12 are branched predominately distributed with Str superfamily members, whereas NC1 has receptors from Srg superfamily. srbc candidates occur at NC7, NC11 exclusively and in NC6 hypotetical protein has a counterpart with a typical GPCR (dct-12). NC11 includes receptors from srbc and srw families. An example study from this cluster association *ie,* srbc-64 and srbc-66 candidates (Fig. S3) are responsible for pheromone activity in *C. elegans* and this illustrates the involvement of serpentine receptors not only in chemoperception, but also proves the fact that all the serpentine receptors are not necessarily non-GPCRs[23,59] few are particular and most are related to potential GPCRs. SS1 carries distinct srj type receptors from Str superfamily with a hypothetical protein.

Cluster 10

Cluster 10 (Fig. S4) includes eight entries of human GPCRs and 26 entries of *C. elegans* GPCRs and is rich in the distribution of srbc type receptors belongs to 'Others' superfamily.[19] Human peptide receptors in this cluster are also distinct (HC1) and do not mix with *C. elegans* GPCRs population further explaining the strong retentionship within species level.[19,45]

Interestingly, the neighbouring clusters NC1–NC2 retain the association of srbc type members only[19] and the adjoining NC3 clade also contains pure set of srw members, while NC4 has pure set of str members. Various type receptors like sri, srt, srg, srh, srd, srt are also associated in NM1 clade further to analyse the diverse properties of serpentine receptors in *C. elegans*.[20]

Cluster 11

Cluster 11 retains 10 human chemokine peptide receptors and 62 *C. elegans* GPCRs (Fig. S4) and exhibits sufficient polyploidy. Though all the entries unite at the root, the association provides NM1 with many more neighbouring members like hypothetical proteins, sre, sra, srh, sri, srw members, typical GPCR members like GAL4, Tag-49 along with 10 related human peptide receptors. Appreciably, two peptide receptors of human (NY1R and NY4R) retained strong association and cluster together within the clade HC1, thereby providing NC2–NC17 as neighbouring clusters (Fig. S4). Also, notably two orthologous pairs are also present in this cluster (refer Table S5). Interestingly, each neighbouring cluster retains its distinct identity with a branching pattern of carrying the same type of receptors belongs to the appropriate superfamily.[40] Thus, NC1, NC2, NC5, NC6, NC15–17 includes the majority of hypothetical proteins and NC3, NC14, NC8, NC10, NC11, NC13, NC14 include members from the largest Str superfamily.[18]

Chemokine Receptors

Cluster 12

Clusters 12 and 13 include candidates from chemokine receptors. Cluster 12 is associated with 16 entries of *C. elegans* GPCRs and 10 human GPCRs (Fig. S4). HC1 accommodates purely with 10 human GPCR entries denoting the evolutionary specificity of human chemokine receptors,[45] Likewise SS1 clade retains purely nematode chemokine receptors of srh type members from Str superfamily. Neighbouring clusters NC1 and NC2 also retain candidates from Str superfamily (srd,str) predominantly, wherein a srt member (NP_507069.1) is associated at the significant *E*-value of 0.86. NC3 is dispersed with a Sra type, hypothetical protein at the bs of 53 and is followed by diverse neighbour members like sri, srbc and Col-40 in NM1.

Cluster 13

In Cluster-13, 15 human GPCRs are associated with 13 *C. elegans* GPCRs. Human GPCRs are distributed in the HC1, HC2 and NM1. Especially, IL-8 A and IL-8B receptors, which are functionally related to calcium storage in human cells,[60] are clustered together at the highest bs of 96 in HC1 clade (Fig. S4). HC2 comprises of receptor for adrenomedullin (ADMR) and Q8NE10. HC3 associates with CCR5 and CCR3 at the bs value of 50 and both of these receptors are implicated in AIDS virology.[61] NM1 includes 9 members of human chemokine receptors with a hypothetical protein from *C. elegans* (NP_504623.1) in the significant *E*-value of 0.003 as neighbour members (NM1) in this cluster. Notably, among all human chemokine receptor, DUFF (the Duffy antigen)[45] is a distantly related receptor but this receptor is observed closer to a hypothetical protein in *C. elegans.*

NC1–NC3 clusters illustrate the distribution of *C. elegans* GPCR entries in neighbouring clusters. Neighbouring cluster NC1 associates with srh members of Str superfamily and NC2 covers hypothetical protein. NC3 cluster includes five members of Str type and a member from srv type receptors. SS1 indicates species-specificity with a member of srxa and a hypothetical protein.

Nucleotide and Lipid Receptors

Clusters from 14–19 associate receptors belonging to nucleotide and lipid type[45] and are majorly activated by negatively charged ligands.[62] These receptors also retain basic residues at the ligand-binding sites and show high sequence diversity owing to the binding of different ligands.[9]

Notably, in this current study, all the species-specific clades observed from Clusters 14 to 19 belong to Str superfamily and observed cluster associations. (SS1–SS8, NC6 in Cluster-14; SS1–SS3, NC2, NC12 in Cluster-15; SS1, SS2 and majority of 'neighbour members' in Cluster-16, SS1 in Cluster-17, SS1 of Cluster-18 and Srd members in SS1 of Cluster-19) explain the unique association of candidate GPCRs mainly from Str superfamily. Such *C. elegans*-only GPCR clusters denote the abundance of this particular superfamily (Str) in the nematode genome.[19,27]

Cluster 14

Cluster-14 retains seven human GPCRs and 64 *C. elegans* GPCRs. Human GPCR members are distributed in HC1 clade and also present in the neighbouring members in NM1. HC1 retains evolutionarily-related human opsin members[9] and four other opsin candidates are dispersed along with the diverse members of *C. elegans* (like srsx, srbc, srx, srt, srg) (Fig. S5) suggesting the evolutionary conservation at the functional level across two taxa for these types of receptors. The neighbouring clusters from NC2–NC5 establish clear composition of srx, srw, srd, srx, respectively wherein NC1 has members from both sre and srg families. NC6 clusters with majority of Str type members (srj and str families) and NM1 observed with srt members.[19,27] SS1–SS5 clusters retain *C. elegans* GPCR members of srh and SS6–SS8 retain sri-type belonging to the large Str superfamily respectively, demonstrating species-specificity.

Cluster 15

Cluster-15 is associated with 18 human GPCRs and 58 *C. elegans* GPCRs. Human GPCR members are distributed into HC1–HC4 along with *C. elegans* candidate GPCRs and 7 other human GPCR entries are dispersed in NM1 (Fig. S5). SS1–SS3 carries clear composition from Str superfamily (srj, sri and str) denoting the species specificity,[19,27] as opposed to NC1–NC13 observed as neighboring clades. NC1 branches at the significant bs of 94 including srv type members of Srg superfamily. NC2–NC4 clade include association from Str superfamily (srh, sri, srm), wherein NC5 includes srj type receptor and a hypothetical protein. NC6 at the bs of 70, associates with a sra-type member with a hypothetical protein. NC7 has str type receptor with hypothetical proteins at the bs value of 62. NC8 includes two members of sri type from Str superfamily and a candidate of srt type from Srg superfamily. NC9 and NC10 associate majorly with srw members. NC11 includes hypothetical proteins at the bs of 54. Notably, NC12 and NC13 have pure association of candidates from sri type and srw type respectively. The NC13 is followed by seven human nucleotide and lipid receptors counterpart with mostly of srw members of *C. elegans* as neighboring members. The human GPCR from HC1–HC4 do not cocluster with *C. elegans* GPCR members, but observed along with neighbour members shows intergenomic clustering.

Cluster 16

In Cluster-16, all the human GPCR members are associated in HC1 clade at the bs of 76,[50] whereas the neighbouring clades, NC1 and NC2 retain pure set of srd GPCRs and a mixture of sra and srxa type, respectively. More *C. elegans* GPCRs such as sri, srh, srx, hypothetical proteins are distributed within Cluster-16 as neighbour members (NM1).[27] SS1 and SS2, however, represent the species-specific clades containing a majority of candidate receptors from Str superfamily (Fig. S5).

Cluster 17

Cluster-17 carries 18 human GPCRs and 30 *C. elegans* GPCRs (Fig. S6). The human GPCR members are dispersed into HC1–HC5. HC1 comprises of cannabinoid receptors (CB1R, CB2R) at the best bs of 93, but the genomes of the protostomian invertebrate like *C. elegans* do not contain CB1R nor FAAH orthologs. This indicates that CB1-like cannabinoid receptors may have evolved after the divergence of deuterostomes.[63–65] Lysophospholipid receptor are clustered in HC2 clade, GP12, GPR3, GPR6 are associated in HC3, Melanocortin receptor observed in HC4 clade, and HC5 carries Sphingosine 1-phosphate receptors.[9] SS1 retains receptors of Srj type and a hypothetical protein and stay distant from the root. The neighbouring clusters were annotated from NC1–NC8. Srv type members associate in NC1, NC2 clade retains srh type,[19,66] NC3 retains pure Srj type members[27] at an average *E*-value of 1.11 (refer Table S4) and NC4 comprises two hypothetical proteins. NC5 clade is of str type[27] with bs of 70. NC6 is of sri type (20) with 65 bs and 0.498 *E*-value, NC7 is with Srh type. NC8 includes a hypothetical protein with a Srsx (Other type) receptor. Many more neighbours with diverse receptor type are observed in NM1. Overall, Cluster-17 covers more number (24 entries) of serpentine types (srh, sri, srj, srg, str) which falls under Str superfamily, and rest of receptors from Srg superfamily became associated and interestingly pure set of clustering belongs to the same family has been observed from NC1–NC7 in this cluster helps to explain the strong intra-genomic retention.

Cluster 18

Cluster-18 consists of 8 Human and 28 *C. elegans* GPCR entries (Fig. S6), which are related to the receptors binding to prostaglandins, prostacyclins and thromboxanes. These receptors bind to ligands which are derivatives of arachidonic acid (AA) and serves as the precursor *via* the cyclooxygenase (COX) pathway. Prostanoids function close to the site of synthesis, and they are deactivated before they are exported into the circulation as inactive metabolites. Prostanoids have essential homeostatic functions in the cytoprotection of gastric mucosa, renal physiology, gestation, and parturition, but they are also implicated in a number of pathological conditions, such as inflammation, cardiovascular disease and cancer. The prostaglandins, prostacyclins and thromboxanes receptors cluster together with a bs of 74 and observed in HC1. The *C. elegans* GPCRs are grouped into NC1–NC5, where NC1 branched at the best bs of 93 with Srt type receptor and hypothetical protein. NC2 retains pure set of receptors from Sra superfamily. NC3 and NC4 also associate with pure set of receptors like srbc, srw from ‘others’ superfamily. Many str type receptors, with a hypothetical protein and a srt type receptor, are associated at NC5. A srxa receptor is observed in NM1. In SS1, receptors from largest Str superfamily with a hypothetical protein are branched with an average *E*-value of 1.178.

Cluster 19

In this cluster, 18 human GPCRs and 93 *C. elegans* GPCRs are associated (Fig. S6), in which human GPCRs are distributed into HC1–HC3. However, around 12 human GPCRs intermix with diverse types of receptors from *C. elegans* (like str, srw, srh, srv, srxa, and hypothetical proteins) suggesting intergenomic association at NM1. Human GPCR clades contain nucleotide and lipid receptor of protease-activated receptors (PAR), psychosine receptors, lysophosphatidylcholine and sphingosylphosphorylcholine. This cluster is highly populated with GPCRs from Str superfamily and ‘Others type’ receptors. NC1–NC24 are denoted as neighbouring clades. Predominantly, receptors like srh, str, srj, sri, srd belong to Str superfamily and are observed at NC1–11, NC13–14, NC18–NC19, NC21–NC22 along with hypothetical proteins and few receptors from ‘Others type’. How-

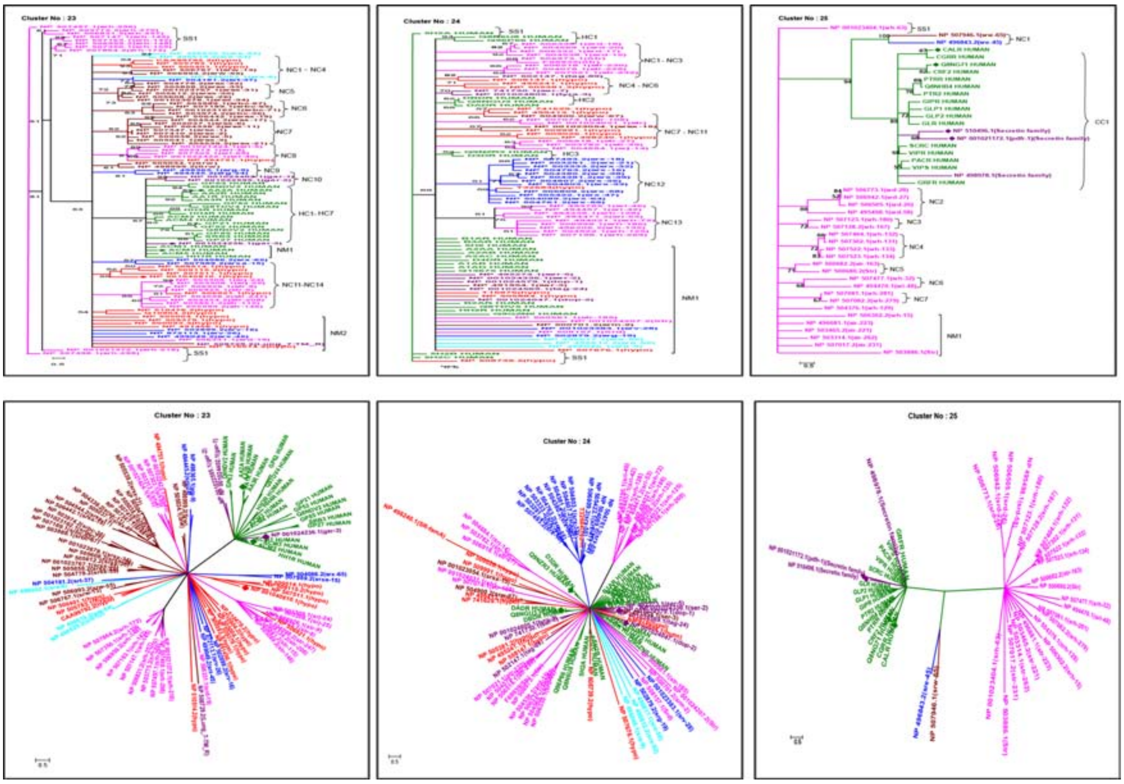

Figure 3. Cross-genome clustering of biogenic amine receptors (Clusters 23, 24) and secretin receptors (Cluster 25) in rectangular (Top) and radial (Bottom) displays.
Notes: Clustering was performed using TREE-PUZZLE 5.1 package and maximum likelihood method. 10,000 quartet puzzling steps were performed and outgroup is finally not shown. Generated newick trees were analyzed using MEGA 4.0.[36] Colour coding is as follows: human GPCRs in green, serpentine receptors of *C. elegans* like Sra superfamily in aqua, Str superfamily in fuchsia/pink, Srg superfamily in blue, Others/Solo type receptors in maroon, typical membrane proteins in purple and hypothetical transmembrane proteins in red. Cluster associations are marked as HC for human GPCRs clade, CC for co-cluster, NC for neighboring cluster, NM for neighbouring members and SS for species-specific clade followed by the number of occurrence in the tree. Orthologous pairs are represented by diamond shaped node markers.

ever, notably all the other neighbouring clades like NC12, NC15–17, NC20, NC23, NC24 are associating with pure set of srw, srbc type receptors. Overall, this cluster is aggregating majorly of Str and 'Others type' superfamily receptors[19] which provide information in connecting these receptor types with biogenic amine receptors of humans.

Biogenic Amine Receptors

Biogenic amine receptors are distributed into five clusters (Cluster-20 to Cluster-24) mainly consisting of trace amine; melatonin; serotonin receptors; histamines, muscarinic acetylcholine, adenosine and histamine; dopamine, octopamine and adrenaline receptors.[9] Intermixing of human and *C. elegans* receptors was observed (Figs. 2 and 3). This suggests biogenic amine receptors have ancient evolutionary origin, as they are observed in invertebrates to higher vertebrates.

Cluster 20

Cluster 20 is represented mainly by trace amine (TA) receptors. Trace amines and their receptors may therefore be useful in treating various neurological and psychiatric disorders.[67–69] and are potentially druggable targets.[70] They form a subfamily of GPCRs related to Norepinephrine (NE), serotonin (5-HT), and dopamine (DA) receptors. Q9P1P5_Hum (GPR58) and Q9P1P4_Hum (GPR57) are closely related to Q96RJ0_Hum (TA1). Similarly, O14804_Hum, a putative neurotransmitter receptor (PNR) is closely related to trace amine (Q969 N4_Hum, Q96RI8_Hum, and Q96RI9_Hum) receptors.[9]

The eight human GPCR entries are branched with very good bs value of 90 at HC1 (Fig. 2) and eight GPCR entries of *C. elegans* GPCRs are distributed as NC1, SS1 and NM1. NC1 carries a pure set of srh type receptors of Str superfamily and a srv type receptor of Srg superfamily observed in NM1. Receptors of Str superfamily (str, srh, srj) at an average *E*-value of 4.23 stay quite distant from the root and has been annotated as a species-specific clade SS1.

Cluster 21

Cluster-21 consists of four human and 49 *C. elegans* GPCRs (Fig. 2). Human GPCRs are distributed into HC1 and CC1. HC1 retains GPCRs belonging to family of melatonin receptors[9] and a candidate GPCR of *C. elegans* NP_493667 (srd 3 type) has been observed in associating with a human GPCR (Q9NQS5_Hum) at a significant *E*-value of 0.01 in CC1 to represent intergenomic association. The other *C. elegans* GPCRs get associated into multiple clades (NC1–NC12). Interestingly NC1, NC4–NC6, NC8–NC10 clades are predominatly associated with receptors of Srg superfamily, whereas NC2, NC7, NC11, NC12 clades carry pure distribution of receptors from largest Str superfamily. Abundant srx type receptors of Srg superfamily are observed with, srbc type receptors and a hypothetical protein at NM1. Overall, despite their association with human biogenic amine receptors, there is a clear branching into various members that belong to Srg and Str superfamilies.

Cluster 22

Cluster-22 consists of 9 human and 29 *C. elegans* GPCR entries (Fig. 2). Human 5-HT$_1$ receptor class comprises of five different receptors, which share 41% to 66% overall sequence identity within themselves and observed in HC1 and HC2. Human GPCRs are distributed into HC1, HC2 and CC1, CC2, NM1 which represent the intermixing of receptors between two taxa. In NM1, 5HT$_{1A}$ of human has an orthologous relationship with NP_497452.1 (ser 4) of *C. elegans* at 1.0e^{-71}. Dopamine 4 receptor of human and tyra-2 of *C. elegans* also got clustered together as ortholog pairs (refer Table S5). *C. elegans* entry, NP_506052.1 (srj type) associates with Q8TDV2 in CC1 at best percentage identity among the rest of all other members as 15.0 at an *E*-value of 0.15 whereas, NP_001024728.1 (ser 1) co-clusters with Q16538 in CC2 and retains the percentage identity as 11.4 at a significant *E*-value of 1.0e^{-50}. Neighbouring clades of NC1–NC7 are present and are mostly retaining pure set of receptor at intragenomic level, wherein receptors from Srg (in NC1), Str (in NC2, NC3), Others (in NC4), Str (in NC5, NC6, NC7) superfamily are observed.

Cluster 23

This cluster contains 22 human GPCRs and 77 *C. elegans* GPCR entries (Fig. 3). G-protein-linked Acetylcholine Receptor family members, like muscarinic acetylcholine, adenosine, histamine and many orphan receptors of human GPCRs, are all clustered in different clades (HC1–HC7) in Cluster 23. Particularly, G-protein-linked Acetylcholine receptor family members, gar-1, gar-2 are observed together as a neighbouring cluster (NC11) and gar-3 of *C. elegans* got associated as neighbouring member (NM1) with human GPCR clades.

NP_001024236.1 (gar-3) has 37.9% of identity with ACM3 human GPCR and retain orthologous relationship and a hypothetical protein NP_001040810.1, retains 7.8% identity with a human GPCR (AA3R) is another notable ortholog, although their functional equivalence is yet to be established.

NC1 is from Sre type belonging to the Srg superfamily, while NC2 associates only hypothetical proteins. Pure set of Srw receptors (Solo Type) are observed in NC3 and NC9, whereas NC4 has receptors of sra and srt types. Notably a pure set of srsx type (Solo Type) receptors are associated in NC5 and NC7. NC6 also retains only Srbc type (Solo Type). NC8 includes majority of srj type and a hypothetical protein.

srg receptors are at NC11 and two hypothetical receptors are at NC12, NC13. NC15 associates with srj, str members of Str superfamily.

Hypothetical proteins, lung even transmembrane receptor and Srx, Srv type receptors are observed at NM2. The species-specific clade (SS1) retains purely srh type receptors of Str superfamily. Overall, Cluster-23 has members from ‘others’/(solo) superfamily as well. Despite the high sequence divergence within human biogenic amine receptors (with seven clades), the fact that many of the *C. elegans* GPCRs associated with this cluster do not co-cluster with the human GPCRs suggests species-specific requirements and lineages of these receptors. Interestingly, this study brings together several uncharacterized and hypothetical proteins to this cluster of biogenic amine receptors.

Cluster 24

Receptors of biogenic amines (dopamine, histamine, octopamine and adrenaline), few serotonergic receptors and many orphan receptors are associated in Cluster 24. This cluster contains 24 GPCRs from human and 68 GPCRs from *C. elegans* (Fig. 3). Human BGA receptors are distributed in HC1–HC3 clades and *C. elegans* GPCRs are noted within clades CC1, NC1–NC13 and NM1 clades. Many biogenic amine receptors of *C. elegans*, like tyra-3, serotonin 5, serotonin 7, serotonin 2, dopamine 1 and dopamine 2 branch near the human biogenic amine receptors, where the percentage identity ranges from 20%–37%. Two pairs of ortholog sets are observed in NM1. NP_001024579.1 (dop-1) has 36.1% of identity with Q8NGU3 and got associated at 1.0e^{-69} and NP_001024047.1 (dop-2) has 33.2% of identity with human D3DR and got associated at 2.00e^{-49}. Notably, neighbouring clades (NC1–NC3, 8, 11 and 13) retain pure set of GPCR members from Str superfamily, whereas the NC4, 5, 7, 9 and 10 clades are predominately associated with hypothetical proteins. Many neighbour members were observed with mixed distribution (receptors belong to srd, srv types from Str superfamily and sre, sra of Sra superfamily).

Class B (secretin) Receptors

Cluster 25

Class B receptors are represented by two clusters (25 and 26) consisting of classical hormone receptors from human and *Drosophila* methuselah (MTH) like proteins.[9] Cluster 25 consists of 17 human and 31 *C. elegans* GPCR entries in which human GPCRs recognise structurally related ligands like polypeptide hormones of 27–141 amino-acid residues (pituitary adenylate cyclase-activating polypeptide (PACAP), secretin, calcitonin, corticotropin-releasing factor (CRF), urocortins, growth-hormone-releasing hormone (GHRH), vasoactive intestinal peptide (VIP), glucagon, glucagon-like peptides (GLP-1, GLP-2) and glucose-dependent insulinotropic polypeptide (GIP) and are related to calcitonin (CALR_Hum) and calcitonin gene-related peptide type 1 receptors (CGRR_Hum). Small accessory proteins (Receptor activity-modifying proteins (RAMPs), interact with these calcitonin receptors and can generate pharmacologically distinct receptors.

Human orphan receptor, Q8NHB4_Hum, is very closely related to PTRR_Hum receptor (that binds to parathyroid hormone and parathyroid hormone-related protein (PTHrP).[9] The human GPCRs within this cluster diversify and in CC1 i GRFR_Hum co-clusters with a receptor belonging to secretin receptor family (NP_498978.1) from *C. elegans.*

The *C. elegans* GPCRs associated with this cluster are distributed as NC1–NC7 (Fig. 3), NM1 and SS1 clades. The NC1 clade has sre and srw receptor types at good bs value of 100 which are associated at average *E*-value of 0.002. The Secretin receptor family of *C. elegans* get association with this cluster. NP_510496.1 (Secretin receptor family) and NP_001021172.1 (pdfr-1) were found as ortholog for CALR_HUMAN (28.5% identity) and CRF2_HUMAN (30.4% identity) at very significant *E*-value. In NC1 NP_498978.1 (Secretin receptor family) retains good percentage identity (28.1) with a human GPCR VIPR_HUMAN. These associations provide examples where co-clustering of orthologs may not happen suggesting functional equivalence even during large sequence variations in this cluster. NC3–NC8 are associated with candidate GPCRs from of srh, srd, str type of Str superfamily, which are distantly related to the human GPCR clades of this cluster and many more neighbour member from same Str superfamily has been observed in NM1. SS1 retains srh and str typereceptors of Str superfamily. In general, Str superfamily members are found more in this cluster.

Cluster 26

As consistent with previous analysis, this particular cluster retains *Drosophila* GPCR members (refer Methods) and due to RPS-BLAST runs, 20 candidate GPCR members from *C. elegans* have been associated. *Drosophila* GPCRs branch into two clades distinctly, whereas *C. elegans* GPCRs are distributed into five sub-clades such as SS1–SS5 (Fig. S7). This cluster could be effective to illustrate the species-specific tendency observed among nematode GPCRs. Methuselah receptors and its paralogs of *Drosophila* solely represent Cluster 26. The *Drosophila* mutant methuselah (MTH) was identified from a screen for single gene mutations that extended average lifespan of an organism and also increased resistance to several forms of stress, including starvation, heat, and oxidative damage. Interestingly, all the clades retain receptors from Str superfamily. SS2 includes entirely srh type receptors from Str superfamily at 69 bs value and many neighbouring members also from Str superfamily with two hypothetical proteins. SS1 also retains candidate receptors from Str superfamily. As mentioned earlier, the predominant occurrence of Str superfamily denotes the abundant availability of str type candidate receptors in nematode genome and is reflects the species-specific tendencies and the limited amount of co-clustering while performing cross-genome phylogeny.

Cell Adhesion Receptors

Cluster 27

This cluster contains 29 human GPCRs and 17 *C. elegans* GPCRs (Fig. S7). A large number of GPCRs belonging to Cell adhesion receptors, characterised by a long extracellular N-terminus and GPCR proteolytic site (GPS) domain, are represented in Cluster 27. Several of these receptors from human have functional domains such as epidermal growth factor (EGF), leucine rich repeat (LRR), hormone-binding domain (HBD) and immunoglobulin (Ig) domains. Hence, they branch into several distantly related clades together with *C. elegans* members and are denoted as HC1 and HC2, CC1, NM1 and SS1. Most of the human GPCRs from this cluster are orphans with no known ligands.[71]

The Q9HAR2 (LEC3 Lectomedin-3) in NM1, has an orthologous relationship with the *C. elegans* gene product NP_001040724.1 (lat-2; *E*-value of 7e^{-71}), *albeit* observed in SS1. lat-2 gene has significant sequence similarity with a paralog—NP_495894.1 (lat-1); lat-1 associates with human Q9HAR2 at an *E*-value of 5e^{-68}. The rest of *C. elegans* GPCRs are branched distantly from the root and has been observed in NC1 with mostly srh, str, sri from Str superfamily. Even though one of the human GPCR (Q8WXG9), co clusters with NC1 clade. This could be due to the large size of Q8WXG9 (6307 amino acids) and can even be viewed as an outlier in this cluster. The clear underrepresentation of *C. elegans* GPCR with cell adhesion receptors, belonging to Cluster 27, is noteworthy.

Class C (glutamate) receptors

Receptors of Class C are divided mainly into four clusters: Clusters 28 to 31. Metabotropic glutamate receptors (MGR), γ-aminobutryic acid (GABA) receptors, calcium-sensing receptors (CASR) and retinoic acid-inducible G-protein-coupled receptors (RAIG) are available as in our previous work.[9]

Cluster 28

Cluster 28 associates with 8 Human and 20 *C. elegans* GPCRs retains a majority of human metabotropic glutamate receptors (MGRs). The primary structure and pharmacology of mGluRs are evolutionarily well-conserved in *Drosophila*, *C. elegans*, and higher mammals.[72]

The human metabotropic glutamate receptor mGluR1, mGluR2, mGluR3, mGluR4, mGluR5, mGluR6, mGluR8, Q8NFS4 (Metabotropic glutamate receptor 7 variant 3) forms a clade and thus inferred as CC1 with the presence of metabotropic glutamate receptor of *C. elegans* (mgl-1, mgl-2). Notably mgl-1 isoform b (CAM33507.1 and gi 125629647) at *E*-value 8.00e^{-87}, mgl-2 (NP_492720.2) at *E*-value 5.00e^{-70} and mgl-3 (NP_741400.1) at *E*-value 3.00e^{-85} has got associated in this cluster (Fig. S7).

Interestingly, mgl-1 of *C. elegans* had been observed as an ortholog of Q8NFS4 (Table S5) and has 51.9% of sequence identity with mGluR2, the mgl-2 (NP_492720.2) has 46.3% of sequence identity with mGluR5 and the mgl-3 (NP_741400.1) has 48.4% of sequence identity with mGluR3 suggesting similar ortholog pairs exist. Specific sequence patterns like SGREL(S/C)Y, TKT, (G/S)RE, MYTTCIIWLAF, NETKFIGFT are well conserved amongst these members (alignment data not shown). NC1–NC6 include Srd, Srh, Str type receptors from Str superfamily. This cluster is illustrative to explain that related *C. elegans* GPCRs, especially from Str superfamily with the glutamate receptor of human do exist and could be identified by our sequence analysis.

Cluster 29

Human calcium-sensing receptor (CASR_Hum-Extracellular calcium-sensing receptor precursor/Parathyroid Cell calcium-sensing receptor) forms Cluster-29 along with a set of 5 orphan receptors and 14 *C. elegans* GPCRs (Fig. S8). Human Calcium-sensing receptor CASR and the orphan receptors (Q8NHZ9, Q8NGV9, 8NGW9 and Q8NGZ7) form a clade with a *C. elegans* GPCR (NP_501400.1). A sweet-taste receptor of 3GCPR/PBP1_GPCR_family_C_like receptor is associated at an *E*-value of 2.00 *E*$^{-26}$ and got branched with Q8NGV9, with 23.46% identity. Thus, the current clustering approach suggests that NP_501400.1 may be a putative ortholog of this extracellular calcium-sensing receptor precursor/

Parathyroid Cell calcium-sensing receptor. NC1 and NC2 clades retain only candidate GPCRs from Str superfamily, in contrast to Sre, Srxa type GPCRs at NC3 and two other Srh type receptors observed at NM1.

Cluster 30

Cluster-30 comprises of four Human GPCRs and 23 *C. elegans* GPCRs (Fig. S8). Human GPCR clade (HC1) contains the human retinoic acid induced GPCRs and orphan GPCR members. *C. elegans* GPCRs form different clades as SS1, NC1, NC2, NM1, NM2 and SS1. Overall, *C. elegans* GPCRs associated with Cluster-30 are from Str superfamily receptors. NC1, NC2 and SS1 clades include *C. elegans* receptors from Str superfamily belonging to srj, srh and str types, respectively; NM1 and NM2 clades also retain receptors of Str superfamily.

Cluster 31

Cluster 31 has four human and eight *C. elegans* GPCRs (Fig. S8). The $GABA_B$ receptors are present in this Cluster. GABAa receptors are members of the ionotropic receptor superfamily which includes alpha-adrenergic and glycine receptors. The four human $GABA_B$ receptors branches with a good bs value of 80 and form the HC1 clade. The NP_741740.1 (gbb-1) receptor was picked as ortholog to GBR1_human by reverse blast best hit procedure with $1.00e^{-48}$ and the NP_493575.2 also gets associated with GBR1_human at a highly significant *E*-value of $4.00e^{-09}$ (Table S5). Both of these *C. elegans* GPCRs get branched in the human clade, HC1. Additionally, NC1 and NM1 clades include receptors from Str and Sra superfamily, respectively and SS1 is associated with Str superfamily

Cluster 32

Frizzed/smoothened receptors

Cluster-32 comprises of receptors with a 200-residue long N-terminus which contains the predicted orthosteric ligand binding site, the cysteine-rich domain (CRD domain) of the receptor[73] which is likely to participate in Wnt ligand binding apart from the GPCR domain. This cluster contains 11 human and 42 *C. elegans* GPCRs (Fig. S8). The Human GPCR (FZD_{1-10}) got associated in CC1. NP_492635.1 (mom5) of *C. elegans* is an ortholog of FZD1 (human GPCR) which got associated with this cluster at an *E*-value of $1.0e^{-75}$ and retains 34.5% identity with human GPCR sequence. Apart from these entries, NP_491028.2 shares 34.5% identity with FZD4 (human GPCR), and NP_503964.2 (cfz2) shares 40.0% identity with FZD5 (human GPCR), got branched in the CC1 itself indicating that there may be a high functional similarity characteristic of close homology between these receptors (Table S5).

The neighbouring clusters were annotated from NC1–NC8. The NC3 clade is of srbc type which forms a clade (with 67 bs value and 0.061 *E*-value) with candidate receptors from Srg superfamily. Neighbouring clusters NC2–NC8 share most of the members from str, srj, srh types from Str superfamily (with <3.0 average *E*-value with good bs values). Interestingly, two hypothetical proteins and receptors of Str superfamily are observed in NM1. Two srh type receptors of *C. elegans* entries are observed in SS1 clade.

Discussion

The selected Human (353)—*C. elegans* (around 1159) GPCRs were associated by RPS-BLAST technique (Tables 1 and 2) to produce 32 cross-genome clusters. In associating nematode GPCRs with 32 profiles of biologically important clusters, grouped under 8 major types of receptors[9] (Figs. 2, 3 and Figs. S2–S8), clustering was done successfully at significant *E*-value thresholds (ranges from 0.001 to 1). 84% and an additional 14% association was observed at the *E*-value thresholds ranges >1 to >5, and very small percentage ie, 2% of association was done by the *E*-value thresholds more than 5 (Fig. S9). Our protocol for associating *C. elegans* sequences to known profile of human GPCR cluster suggest that there has been high representation of *C. elegans* GPCRs (except in the cases of chemokine receptor cluster (Cluster 12, 13) and cell adhesion receptors (Cluster 27) (refer Table 1). The distribution of serpentine receptors at superfamily level, associated hypothetical proteins, and orthologs (given in Fig. S9A), also a trial study with known association (Table 3) clearly supports the RPS-blast clustering technique, and particularly the association of orthologs (refer Table S5) at significant *E*-values is a clear evidence for clustering GPCR sequence across genome by using Profile based sequence associating method—RPS blast technique.

Since the significant *E*-values play a major role and act as a preliminary reference for sequence comparision, the current method helps to connect the reported hypothetical proteins (refer Table S6) with the associated known receptor types at sequence level further to compare for the functional relationship.

The resultant 32 enriched clusters and their phylogenies were analysed and discussed in detail for the distribution in cross-genome studies. For this, we have introduced some terms, like human GPCR clade [HC] Co-clusters [CC], Neighbour Clades [NC], Neighbour Members [NM], Species specific members [SS] to follow the branching patterns in the dendrogram of different clusters (Figs. 2A–C, 2, 3 and Figs. S2–S8).

Further, the association of GPCRs in two genomes by our sequence analysis suggests that we can capture remote homology from 12% to 20% as average cluster identity (refer Supplementary File) and can include highly related (co-clusters, orthologs), related (neighbour clusters) and distantly related (neighbour members). A broad spectrum of sequence relationships between human and *C. elegans* GPCRs could be seen: for example, there is intermixing in biogenic amine receptors (Clusters 20 to 24), sufficient polyploidy amongst members in a cluster (example as in Clusters 6 and 11), not sufficient intermixing (as observed in Clusters 10 and 26) and strong species-specific tendencies (example as noticed for nucleotide and lipid receptors (Clusters 14 to 19).

The identification of orthologs and putative orthologs (Table S5, for example as in Clusters 1, 5, 8 and 21) among GPCRs of the two genomes helps to correlate the evolutionary integrity between the two genomes. Interestingly, in many instances, we could observe the orthologs co-clustered within the same clade (example: in cluster 1 the npr-9 is ortholog to GALR in peptide receptor) validating our associations and clustering techniques. In this study, unannotated and hypothetical proteins (Table S6) are associated with GPCR clusters at statistically significant *E*-values provoking their function to be interrogated by experiments (for example as in Clusters 3–5, 8, 11, 16–17, 23 and 32). No intermixing of sequence groups across individual genomes (species-specific (SS) clades) and in Human GPCR (HC) clades) have also been noticed in some instances (for example, Clusters 10 and 26). Nematode-specific serpentine receptors, with their Pfam domain knowledge, GO annotation (Table S1) are helpful to understand the species-specific repertoire of GPCRs for the Sra, Str, Srg and others/Solo type of superfamilies. The motifs belong to these receptors and the distribution of these receptors among biologically important eight subtypes of human GPCRs (Table S7) further helps to address the nematode specificity and to provide guidelines to understand species-specific sequence properties.

Our recent publication on identifying conserved motifs in the aligned set of cross-genome GPCR clusters are biologically useful to connect the conservation at sequence level next to structure then to functional benefit.[73,74]

Overall, our cross–genome study, using sequence search and clustering strategy, uncovers information on putative orthologs, function annotation of novel genes and functionally important sequences in two genomes for further practical applications.

Conclusions

Reported associations in the current study for selected human and *C. elegans* GPCRs provide information at a preliminary level of understanding the cross-genome clustering and possible related sequences across taxa.

Though the RPS-blast is a sensitive approach in associating the queries to the profile not with reference to the sequence identity (independent of sequence identity), and lack of input of all the available/possible receptor type in profiles, the current approach is effective in connecting the remote homologs to the given representative profiles. In future, by observing the fine grained analytical approaches like identifying conserved domains, motifs will provide clear understanding about the established associations.

The current study suggests an approach to perform cross-genome analysis of particular protein families by employing simple clustering algorithms. We suggest objective methods of studying such cross-genome phylogenies to recognise highly conserved intergenomic co-clusters and species-specific intragenomic clades.

The study also reports for the orthologous pairs, and connecting Olfactory receptors (ORs), pheromone–like receptors for the correct association with the respective Str rich GPCR clusters. For instance, the Odr-10, is associated in the cluster 5 (peptide receptor), but established the cluster association only with Str type GPCRs and notably observed in SS3 clade (refer details in Cluster 5 in results). The other interesting evidence is the recently reported[69] serpentine receptors like Srg 36, 37 were given chance to pick up the closest representative profile. Separately, RPS-BLAST reports the association at cluster 25 (Secretein type receptors) and cluster 24 (Biogeneic amine receptors) respectively with the significant *E*-value cut-off of 3 e-04 and 0.2 (refer Table 3 in additional information). This association is particularly helpful in explaining the trend of association in *C. elegans* GPCRs to the related serpentine receptor super families, ie, the said (Srg 36, 37) receptors clustered in the clade with candidates of Str superfamily, also notably the associated clusters are rich with Str-type receptors (refer Cluster 24, 25 for details).

The discussed case studies are support the effectiveness of RPS-BLAST for associating the related sequences to its nearest superfamilies (case studies on srg 36, 37), though the lack of profiles for OR-like GPCRs; these receptors tend to be in separate clade to retain species specific property with the correcpoding Str superfamily candidate GPCRs and picking up the closest receptor–profile.

Beside evolutionary pressures, certain GPCRs retain conservation and in parallel, exhibit some level of integrity [refer CC observations at cluster 1, 4, 5, 7, 8 in peptide receptors, cluster 22 in Biogenic amine receptors, cluster 25 in secretine type receptor, cluster 32 in frizzled type receptors], also the observed unannotated GPCRs in the association can be further analysed for the functional relevance and to implement it for practical benefit.

Author Contributions

Conceived and designed the experiments: RS. Analysed the data: BN, SK. Wrote the first draft of the manuscript: BN, SK. Contributed to the writing of the manuscript: RS, VB. Agree with manuscript results and conclusions: BN, SK, VB, RS. Jointly developed the structure and arguments for the paper: BN, SK, VB, RS. Made critical revisions and approved final version: RS. All authors reviewed and approved of the final manuscript.

Funding

This work is supported by an Indo-Japan Grant funded by Department of Biotechnology,India and AIST, Japan.

Competing Interests

There are no conflicts of interest to declare.

Acknowledgements

We sincerely express our gratitude to NCBS (TIFR) for support and infrastructural facilities.

Disclosures and Ethics

As a requirement of publication author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

References

1. Marinissen MJ, Gutkind JS. G-protein-coupled receptors and signaling networks: emerging paradigms. *Trends Pharmacol Sci*. 2001;22:368–76.
2. Dianne M. Perez. From Plants to Man: The GPCR "Tree of Life". *Mol Pharmacol*. 2005;67:1383–4.
3. Sabine Schlyer, Richard Horuk. I want a new drug: G-protein-coupled receptors in drug development. *Drug Discovery Today*. 2006;11:481–93.
4. Steven C Prinster, Chris Hague Randy A Hall. Heterodimerization of G protein-coupled receptors. Specificity and functional significance. *Pharmacological Reviews*. 2005;57:3289–98.
5. Iva Greenwald. Introduction to signal transduction. *Worm Book*. 2005.
6. Kuwabara PE, neil NO'. The use of functional genomics in *C. elegans* for studying human development and disease. *J Inherit Metab Dis*. 2001;24: 127–38.
7. Fredriksson R, Schioth HB. The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Mol Pharmacol*. 2005;67:1414–25.
8. *C. elegans* Sequencing Consortium: Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*. 2012;282:1998. PMID: 9851916
9. Raghu Prasad Rao Metpally, Ramanathan Sowdhamini. Cross genome phylogenetic analysis of human and *Drosophila* G protein-coupled receptors: application to functional annotation of orphan receptors. *BMC Genomics*. 2005;6(106):1–20.

10. Brenner S. The genetics of *Caenorhabditis elegans*. *Genetics*. 1974;77: 71–94.
11. Khader Shameer P, Nagarajan K, Gaurav, Sowdhamini R. 3PFDB-database of Best Representative PSSM Profiles (BRPs) of Protein Families generated using a novel data mining approach *Bio Data Mining*. 2009;2:8.
12. Gowri VS, Khader Shameer, Chilamakuri Chandra Sekhar Reddy, Prashant Shingate, Ramanathan Sowdhamini. A Sequence data mining protocol to identify best representativesequence for protein domain families. Proceedings of 2010 IEEE International Conference on Data Mining Workshops (ICDMW 2010). In press.
13. Marchler-Bauer A, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res*. Jan 2011;39(Database issue):D225–9. Epub Nov 24, 2010. PubMed PMID: 21109532; PubMed Central PMCID: PMC3013737.
14. Maxwell CK Leung, Phillip L Williams, Alexandre Benedetto, et al. *Caenorhabditis elegans*: an emerging model in biomedical and environmental toxicology. *Toxicological Sciences*. 2008;106(1):5–28.
15. Graul RC, Sadee W. Evolutionary relationships among G protein-coupled receptors using a clustered database approach. *AAPS Pharm Sci*. 2001; 3(2):Article.
16. Lee S, Horn V, Julien E, et al. Epigenetic regulation of histone H3 serine 10 phosphorylation status by HCF-1 proteins in *C. elegans* and mammalian cells. *PLoS One*. Nov 28, 2007;2(11):e1213. PubMed PMID: 18043729; PubMed Central PMCID: PMC2082077.
17. Cui M, Kim EB, Han M. Diverse chromatin remodeling genes antagonize the rb-involved synmuv pathways in *C. elegans*. *PLoS Genet*. 2006; 2(5):e74.
18. Maido Remm, Erik Sonnhammer. Classification of Transmembrane protein families in the *Caenorhabditis elegans* genome and identification of human orthologs, *Genome Res*. 2000;10:1679–89.
19. Hugh M Robertson. Two large families of chemoreceptor genes in the nematodes caenorhabditis elegans and caenorhabditis briggsae reveal extensive gene duplication, diversification, movement, and intron loss. *Genome Res*. 1998;8:449–63.
20. Troemel ER, Chou JH, Dwyer ND, Colbert HA, Bargmann CI. Divergent seven transmembrane receptors are candidate chemosensory receptors in *C. elegans*. *Cell*. 1995;83:207–21.
21. Nansheng Chen, Shraddha Pai, Zhongying Zhao, Allan Mah, Rebecca Newbury, Robert C Johnsen. Identification of a nematode chemosensory gene family. *PNAS*. 2005;102(1):146–51.
22. Emily R Troemel. Chemosensory signaling in *C. elegans*. *Bio Essays*. 1999; 21(12):1011–20.
23. Robertson HM, Thomas JH. The putative chemoreceptor families of *C. elegans* (Jan 6, 2006), *WormBook*, ed. The *C. elegans* Research Community, WormBook. doi/10.1895/wormbook.1.66.1.
24. Melkman T, Sengupta P. The worm's sense of smell Development of functional diversity in the chemosensory system of Caenorhabditis elegans. *Dev Biol*. Jan 15, 2004;265(2):302–19.
25. McGrath, Patrick T, Xu, et al. Parallel evolution of domesticated Caenorhabditis species targets pheromone receptor genes. *Nature*. Aug 17, 2011:1476–4687. http://dx.doi.org/10.1038/nature10378.
26. Kim K, Sato K, Shibuya M, et al. Two chemoreceptors mediate developmental effects of dauer pheromone in *C. elegans*. *Science*. Nov 13, 2009; 326(5955):994–8. Published online Oct 1, 2009.
27. Robertson HM. Updating the Str and srj (stl) families of chemoreceptors in Caenorhabditis nematodes reveals frequent gene movement within and between chromosomes. *Chem Senses*. Feb 2001;26(2):151–9.
28. Woollard A. Gene duplications and genetic redundancy in *C. elegans* (Jun 25, 2005), *WormBook*, ed. The *C. elegans* Research Community, WormBook. doi/10.1895/wormbook.1.2.1, http://www.wormbook.org.
29. Ono Y, Fujibuchi W, Suwa M. Automatic gene collection system for genome-scale overview of G-protein coupled receptors in eukaryotes. *Gene*. 2005;364:63–73.
30. Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*. 1998;14:378–9.
31. Tusnády GE, Simon I. Principles governing amino acid composition of integral membrane proteins: applications to topology prediction. *J Mol Biol*. 1998;283:489–506.
32. Altschul SF, Madden TL, Schaffer AA. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. 1997;25(17):3389–402.
33. Pirovano W, FeenStra KA, Heringa J. PRALINETM: a Strategy for improved multiple alignment of transmembrane proteins. *Bioinformatics*. 2008;24(2):492–7.
34. Schmidt HA, Strimmer K, Vingron M, von Haeseler A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*. 2002;18:502–4.
35. Strimmer K, von Haeseler A. Quartet puzzling. A quartet maximum likelihood method for recon. Structing tree topologies. *Mol Biol Evol*. 1996;13: 964–9.
36. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*. 2007;24:1596–9.
37. Seminara SB, Messager S, Chatzidaki EE, et al. The GPR54 gene as a regulator of puberty. *N Engl J Med*. 2003;349:1614–27.
38. de Roux N, Genin E, Carel JC, Matsuda F, Chaussain JL, Milgrom E. Hypogonadotropic hypogonadism due to loss of function of the KiSS1-derived peptide receptor GPR54. *Proc Natl Acad Sci U S A*. 2003;100:10972–6.
39. Semple RK, Achermann JC, Ellery J, et al. Two novel missense mutations in G protein-coupled receptor 54 in a patient with hypogonadotropic hypogonadism. *J Clin Endocrinol Metab*. 2005;90:1849–55.
40. Teles MG, Bianco SDC, Brito VN, et al. A GPR54-activating mutation in a patient with central precocious puberty. *N Engl J Med*. 2008;358:709–15.
41. Matsumoto M, Kamohara M, Sugimoto T, et al. The novel G-protein coupled receptor SALPR shares sequence similarity with somatostatin and angiotensin receptors. *Gene*. 2000;248:183–9.
42. http://en.wikipedia.org/wiki/Rhodopsin-like_receptors#Function.
43. Heine U, Blumenthal T. Characterization of regions of the *Caenorhabditis elegans* X chromosome containing vitellogenin genes. PMID: 3735423.
44. *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*. Dec 11, 1998; 282(5396):2012–8. Review. Erratum in: Science. Jan 1, 1999;283(5398):35. *Science*. Mar 26, 1999;283(5410):2103. *Science*. Sep 3, 1999;285(5433): 1493. PubMed PMID: 9851916.
45. 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*. Oct 17, 2002;3(11):RESEARCH0063. Epub Oct 17, 2002. PubMed PMID: 12429062; PubMed Central PMCID: PMC133447.
46. Thomas JH, Robertson HM. The Caenorhabditis chemoreceptor gene families. *BMC Biol*. Oct 6, 2008;6:42. PubMed PMID: 18837995; PubMed Central PMCID: PMC2576165.
47. Thomas R, Chen J, Roudier MM, Vessella RL, Lantry LE, Nunn AD. In vitro binding evaluation of 177 Lu-AMBA, a novel 177 Lu-labeled GRP-R agonist for systemic radiotherapy in human tissues. *Clin Exp Metastasis*. 2009;26(2):105–9. Epub Oct 31, 2008. PMID: 18975117.
48. Ohki-Hamazaki H, Watase K, Yamamoto K, et al. Mice lacking bombesin receptor subtype-3 develop metabolic defects and obesity. *Nature*. Nov 13, 1997;390(6656):165–9.
49. Villeneuve A, Gignac S, Provencher PH. Glucocorticoids decrease endothelin-A- and -B-receptor expression in the kidney. *J Cardiovasc Pharmacol*. Nov 2000;36(5 Suppl 1):S238–4.
50. http://en.wikipedia.org/wiki/Rhodopsin-like_receptors.
51. Piali Sengupta, et al. *odr*-10 encodes a seven transmembrane domain olfactory receptor required for responses to the odorant diacetyl. *Cell*. Mar 22, 1996;84:875–87.
52. Vadakkadath Meethal S, Gallego MJ, Haasl RJ, Petras SJ 3rd, Sgro JY, Atwood CS. Identification of a gonadotropin-releasing hormone receptor orthologue in Caenorhabditis elegans. *BMC Evol Biol*. Nov 29, 2006;6:103.
53. Sonia Terrillon, Claude Barberis, Michel Bouvier. Heterodimerization of V1a and V2 vasopressin receptors determines the interaction with beta-arrestin and their trafficking patterns. *PNAS*. 2004:1548–53.

54. Daniel J, Scott, Sharon Layfield, et al. Bathgate, Characterization of novel splice variants of LGR7 and LGR8 reveals that receptor signaling is mediated by their unique low density lipoprotein class a modules. *Journal of Biological Chemistry*. 281:34942–54.
55. Cho S, Rogers KW, Fay DS. The *C. elegans* glycopeptide hormone receptor ortholog, FSHR-1, regulates germline differentiation and survival. *Curr Biol*. Feb 6, 2007;17(3):203–12.
56. Devanjan Sikder, Thomas Kodadek. The neurohormone orexin stimulates hypoxia-inducible factor-1 activity. *Genes and Dev*. 2007;21:2995–3005.
57. Ulrich CD, Ferber I, Holicky E, Hadac E, Buell G, Miller LJ. Molecular cloning and functional expression of the human gallbladder cholecystokinin A receptor. *Biochem Biophys Res Commun*. 1993;193:204–11.
58. Didier Rognan. Development and virtual screening of target libraries. *Journal of Physiology-Paris*. Mar–May 2006;99(2–3):232–44.
59. Murphy PM, Tiffany HL. Cloning of complementary DNA encoding a functional human interleukin-8 receptor. *Science*. 13, 1991;253(5025):1280–3.
60. Michelle S Teng, Martijn PJ Dekkers, Bee Ling Ng, et al. Expression of mammalian GPCRs in *C. elegans* generates novel behavioural responses to human ligands. *BMC Biology*. 2006;4:22.
61. Stefano Costanzi, Susanne Neumann Marvin C Gershengorn. Seven transmembrane-spanning receptors for free fatty acids as therapeutic targets for diabetes mellitus: pharmacological, phylogenetic, and drug discovery aspects. *The Journal of Biological Chemistry*. 283:16269–73.
62. Montero C, Campillo NE, Goya P, Paez JA. Homology models of the cannabinoid CB1 and CB2 receptors. A docking analysis study. *Eur J Med Chem*. 2005;40:75–83. doi: 10.1016/j.ejmech.2004.10.002.
63. Elphick MR, Egertova M. The neurobiology and evolution of cannabinoid signaling. *Philos Trans R Soc Lond B Biol Sci*. 2001;356:381–408.
64. Elphick MR. An invertebrate *G*-protein coupled receptor is a chimeric cannabinoid/melanocortin receptor. *Brain Res*. 1998;780:170–6.
65. McCarroll SA, Li H, Bargmann C. Identification of transcriptional regulatory elements in chemosensory receptor genes by probabilistic segmentation. *Curr Biol*. 2005;15:347–52.
66. Thomas JH. Analysis of Homologous gene clusters in *C. elegans* reveals striking regional cluster domains. *Genetics*. 2006. 10.1534/geneticS104.040030.
67. Branchek TA, Blackburn TP. Trace amine receptors as targets for novel therapeutics: legend, myth and fact. *Curr Opin Pharmacol*. 2003;3:90–7.
68. Berry MD. The potential of trace amines and their receptors for treating neurological and psychiatric diseases. *Rev Recent Clin Trials*. 2007;2:3–19.
69. Davenport AP. Peptide and trace amine orphan receptors: prospects for new therapeutic targets. *Curr Opin Pharmacol*. 2003;3:127–34.
70. Foord SM, Jupe S, Holbrook J. Bioinformatics and type II G-protein-coupled receptors. *Biochem Soc Trans*. 2002;30:473–9.
71. Hideo Taniura, Noriko Sanada, Nobuyuki Kuramoto, Yukio Yoneda. A metabotropic glutamate receptor family gene in dictyostelium discoideum. *JBC Papers in Press*. Mar 9, 2006. doi:10.1074/jbc.M512723200.
72. Adams MD, et al. The genome sequence of *Drosophila melanogaster*. *Science*. Mar 24, 2000;287(5461):2185–95. PubMed PMID:10731132.
73. Balasubramanian Nagarathnam, Shankar Kannan, Vardhan Dharnidharka, Veluchamy Balakrishnan, Govindaraju Archunan, Ramanathan Sowdhamini. Insights from the analysis of conserved motifs and permitted amino acid exchanges in the human, fly and worm GPCR clusters. *Bioinformation*. 2011;7(1):15–20.
74. Balasubramanian Nagarathnam, Kannan Sankar, Varadhan Dharnidharka, Veluchamy Balakrishnan, Govindaraju Archunan, Ramanathan Sowdhamini. TM-MOTIF: an alignment viewer to annotate predicted transmembrane helices and conserved motifs in aligned set of sequences. *Bioinformation*. 2011.

Supplementary Data

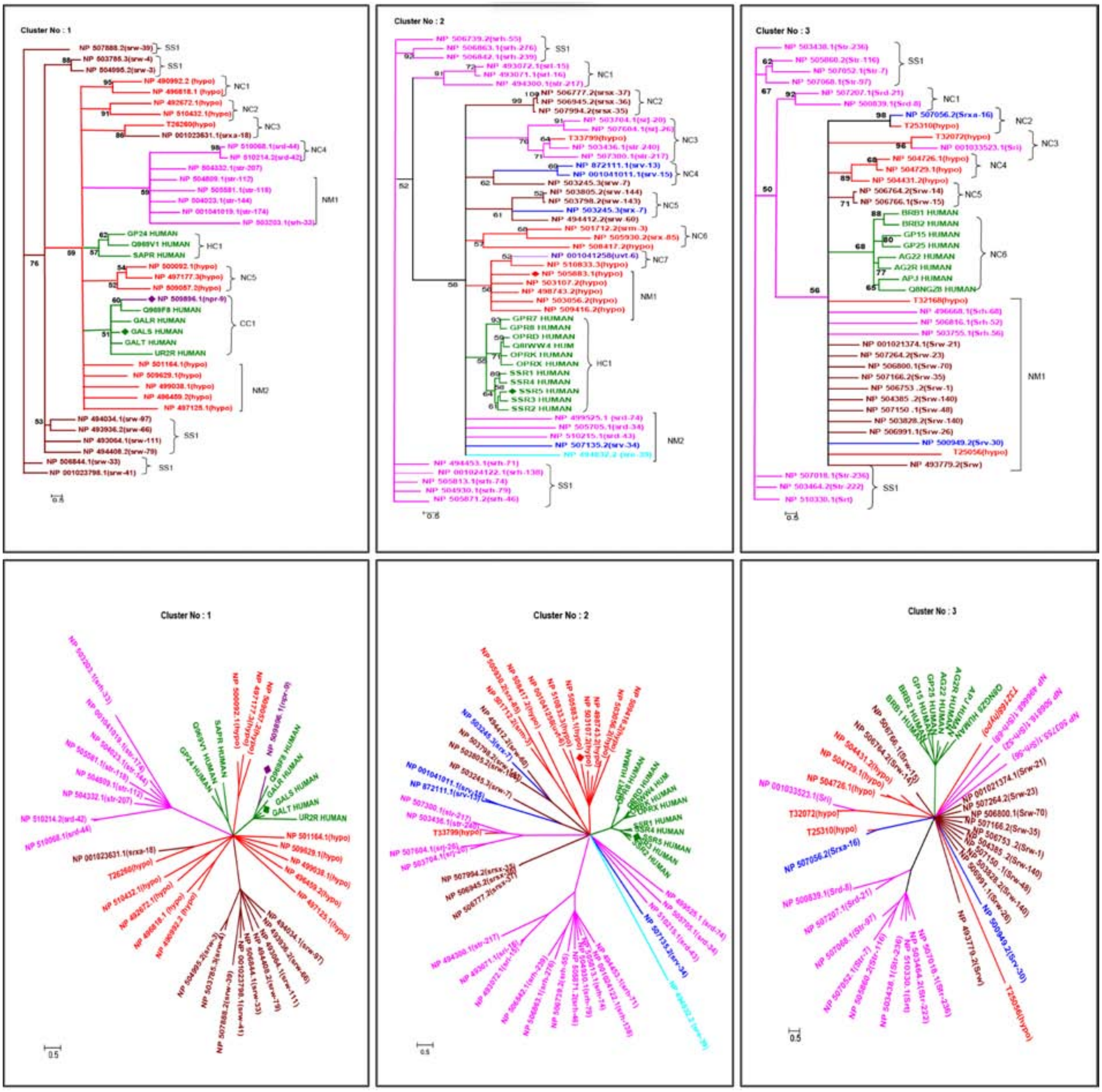

Figure S1. Phylogenetic trees of peptide receptors (Clusters 1–3) in rectangular (Top) and radial (Bottom) displays.

Notes: Clustering was performed using TREE-PUZZLE 5.1 package and maximum likelihood method. 10,000 quartet puzzling steps were performed and outgroup is finally not shown. Generated newick trees were analyzed using MEGA 4.0.[36] Colour coding is as follows: human GPCRs in green, serpentine receptors of *C. elegans* like Sra superfamily in aqua, Str superfamily in fuchsia/pink, Srg superfamily in blue, Others/Solo type receptors in maroon, typical membrane proteins in purple and hypothetical transmembrane proteins in red. Cluster associations are marked as HC for human GPCRs clade, CC for co-cluster, NC for neighboring cluster, NM for neighbouring members and SS for species-specific clade followed by the number of occurrence in the tree. Orthologous pairs are represented by diamond shaped node markers.

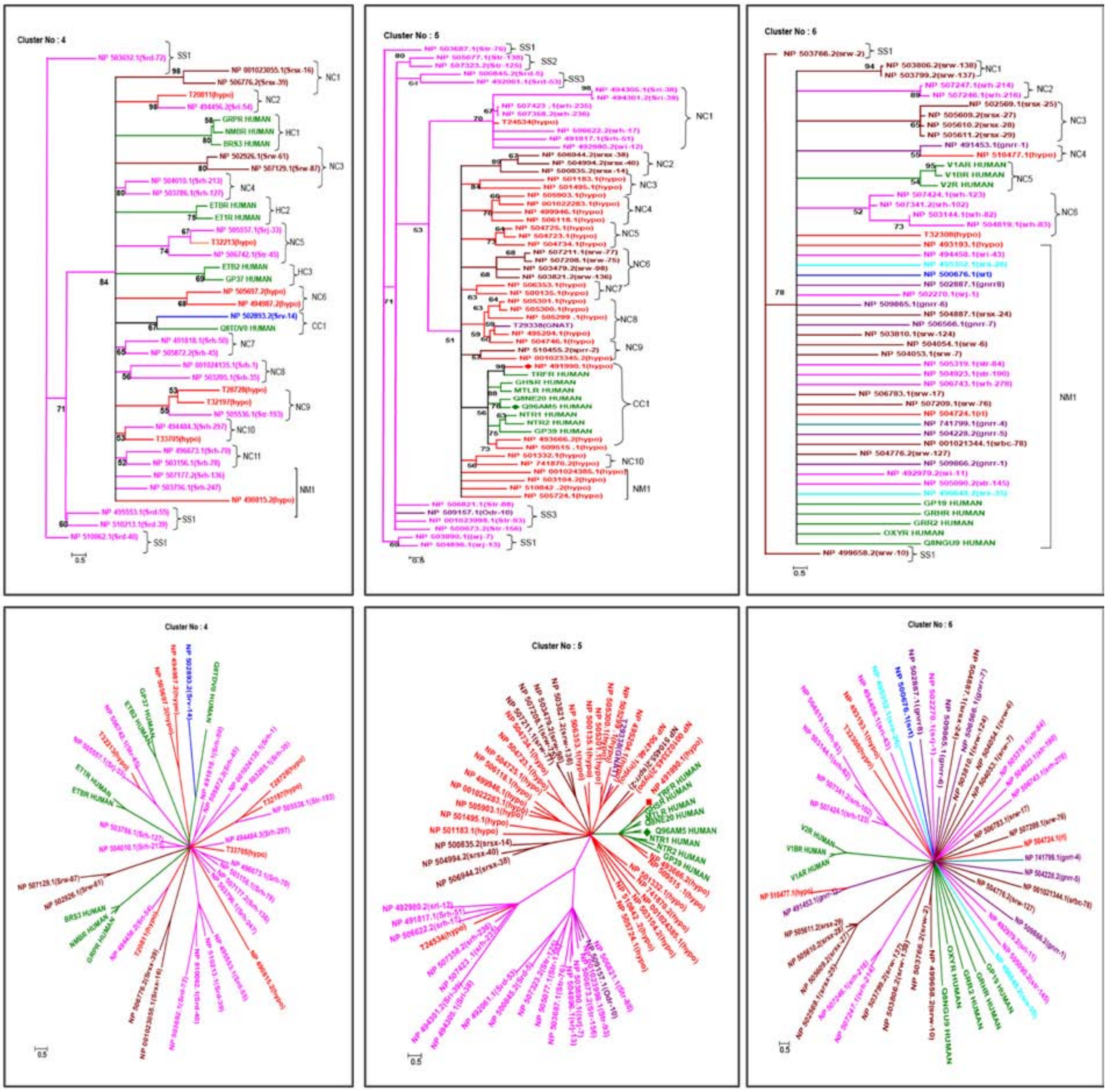

Figure S2. Same as Figure S1 but for peptide receptors (Clusters 4–6).

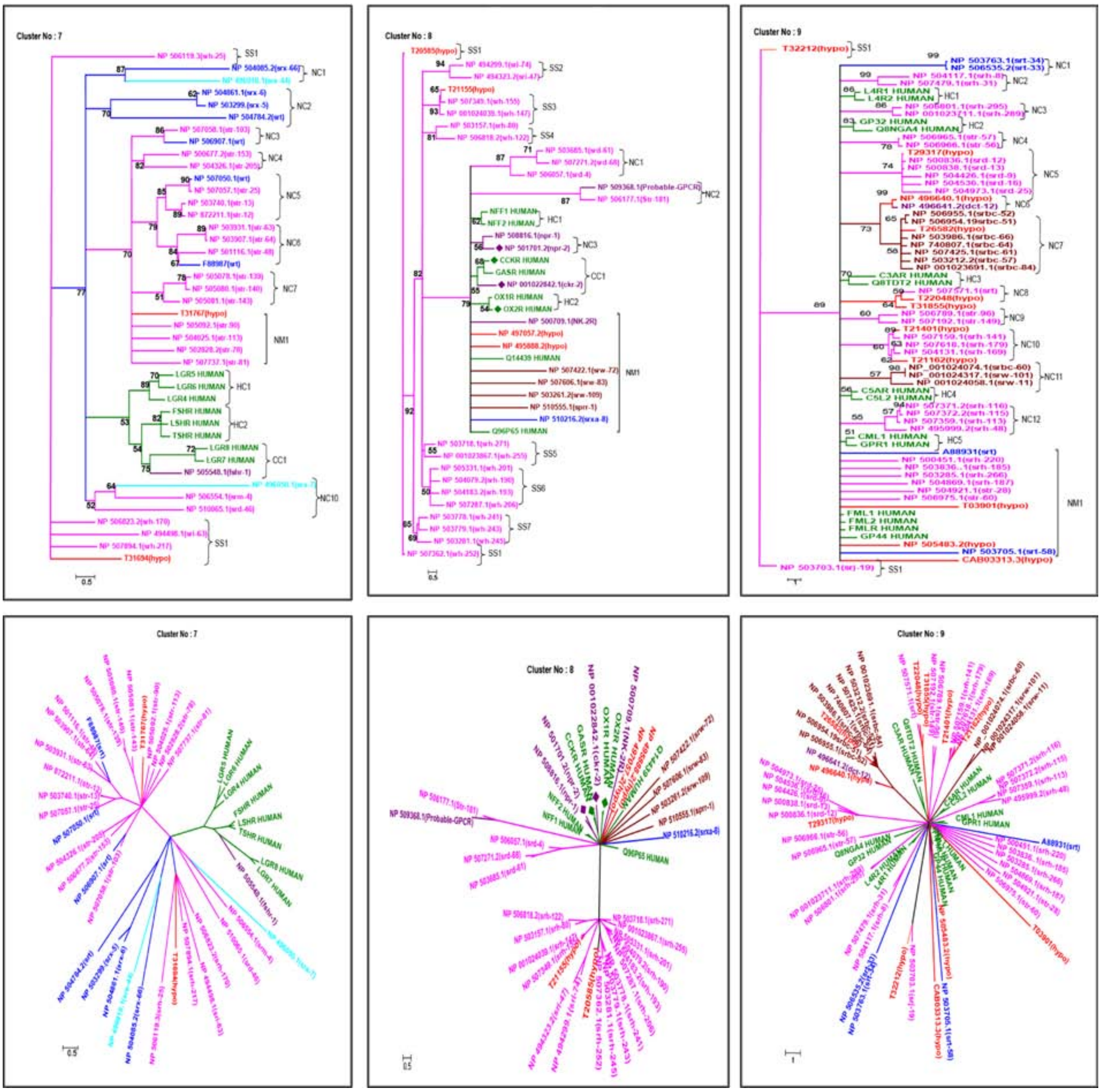

Figure S3. Same as Figure S1 but for peptide receptors (Clusters 7–9).

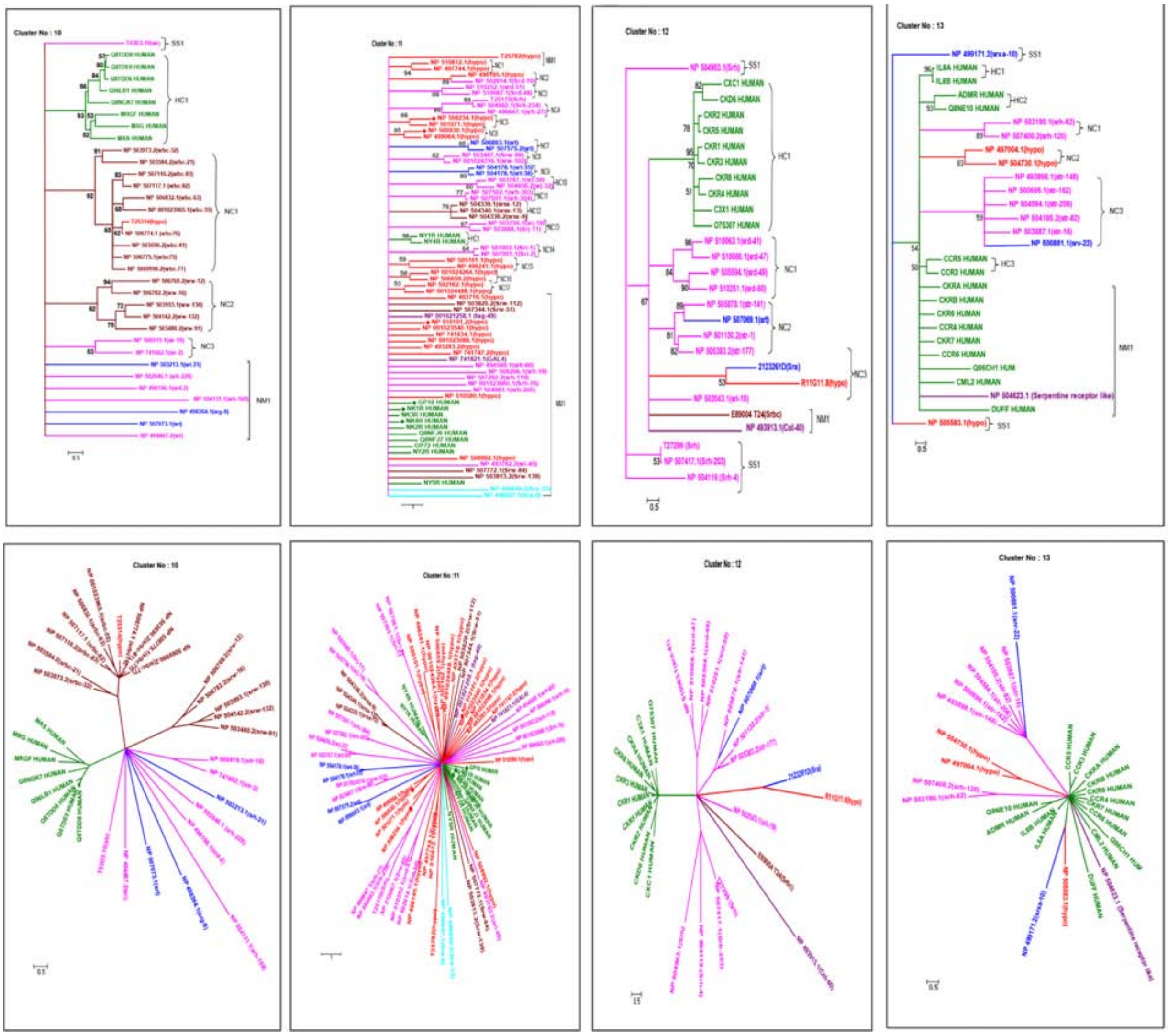

Figure S4. Same as Figure S1 but for peptide receptors (Clusters 10, 11), chemokine receptors (12, 13).

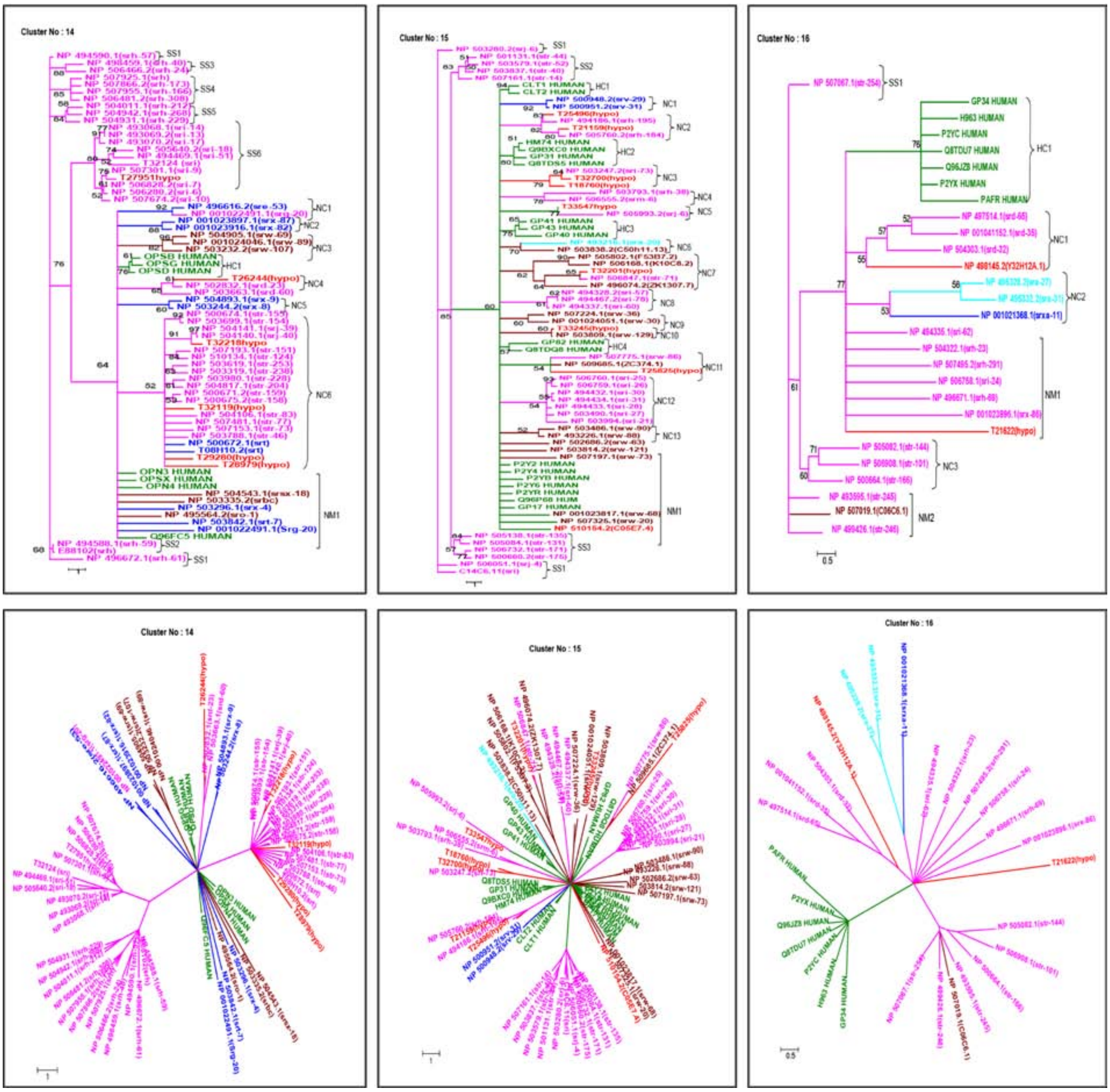

Figure S5. Same as Figure S1 but for nucleotide and lipid receptors (Clusters 14–16).

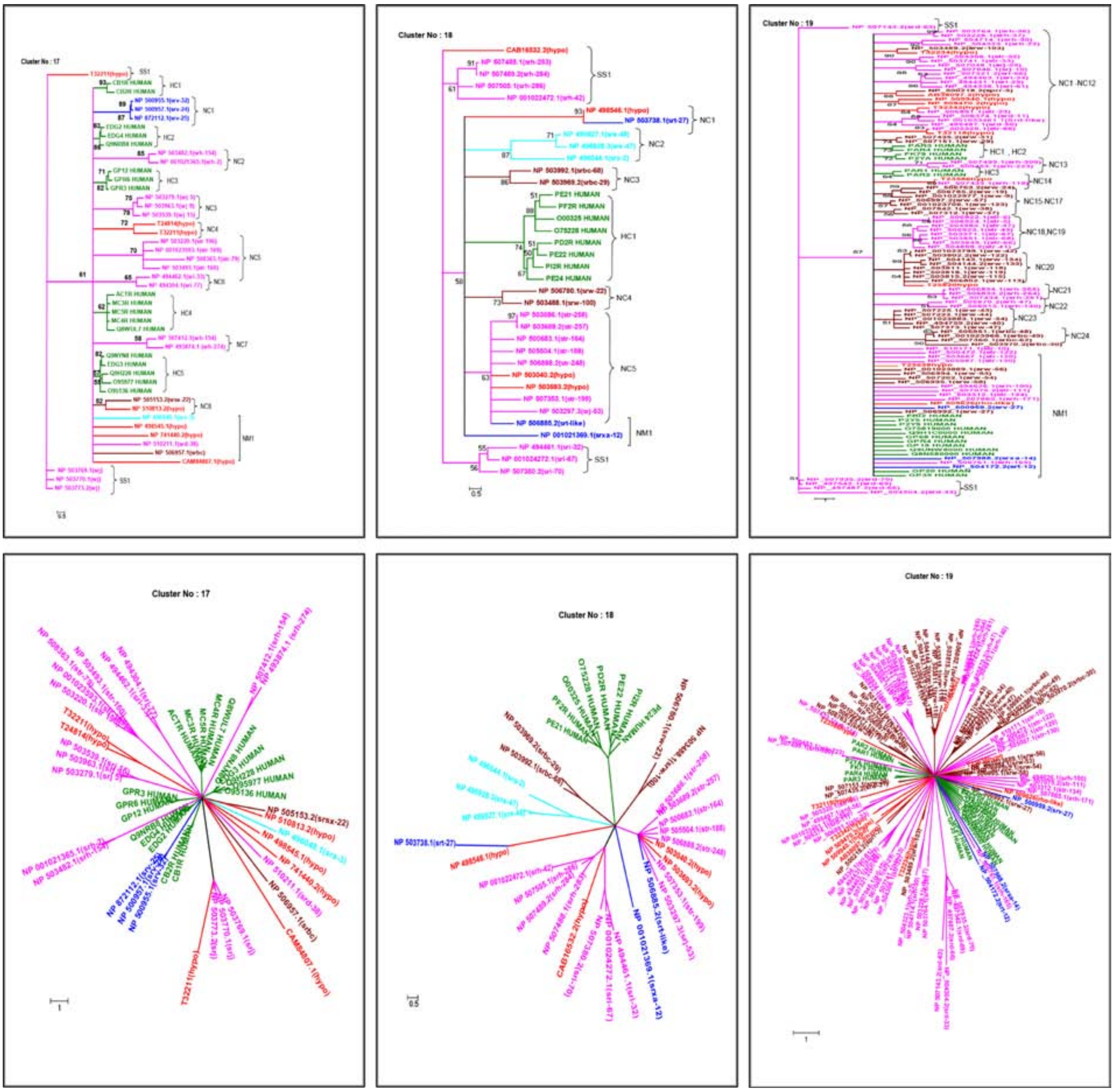

Figure S6. Same as Figure S1 but for nucleotide and lipid receptors (Clusters 17–19).

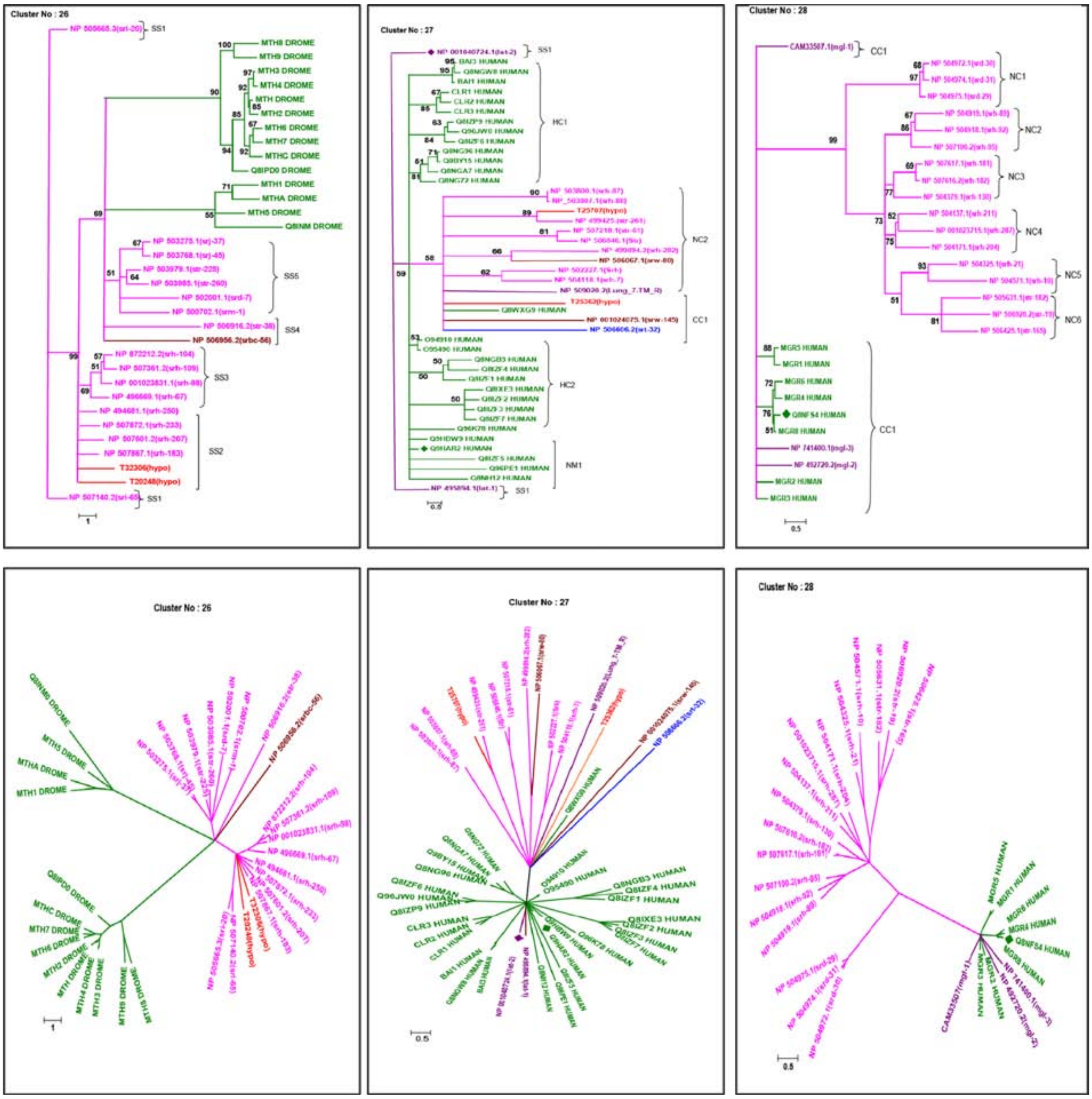

Figure S7. Same as Figure S1 but for secretin receptors (Clusters 26), cell adhesion receptors (Cluster 27), glutamate receptors (Cluster 28).

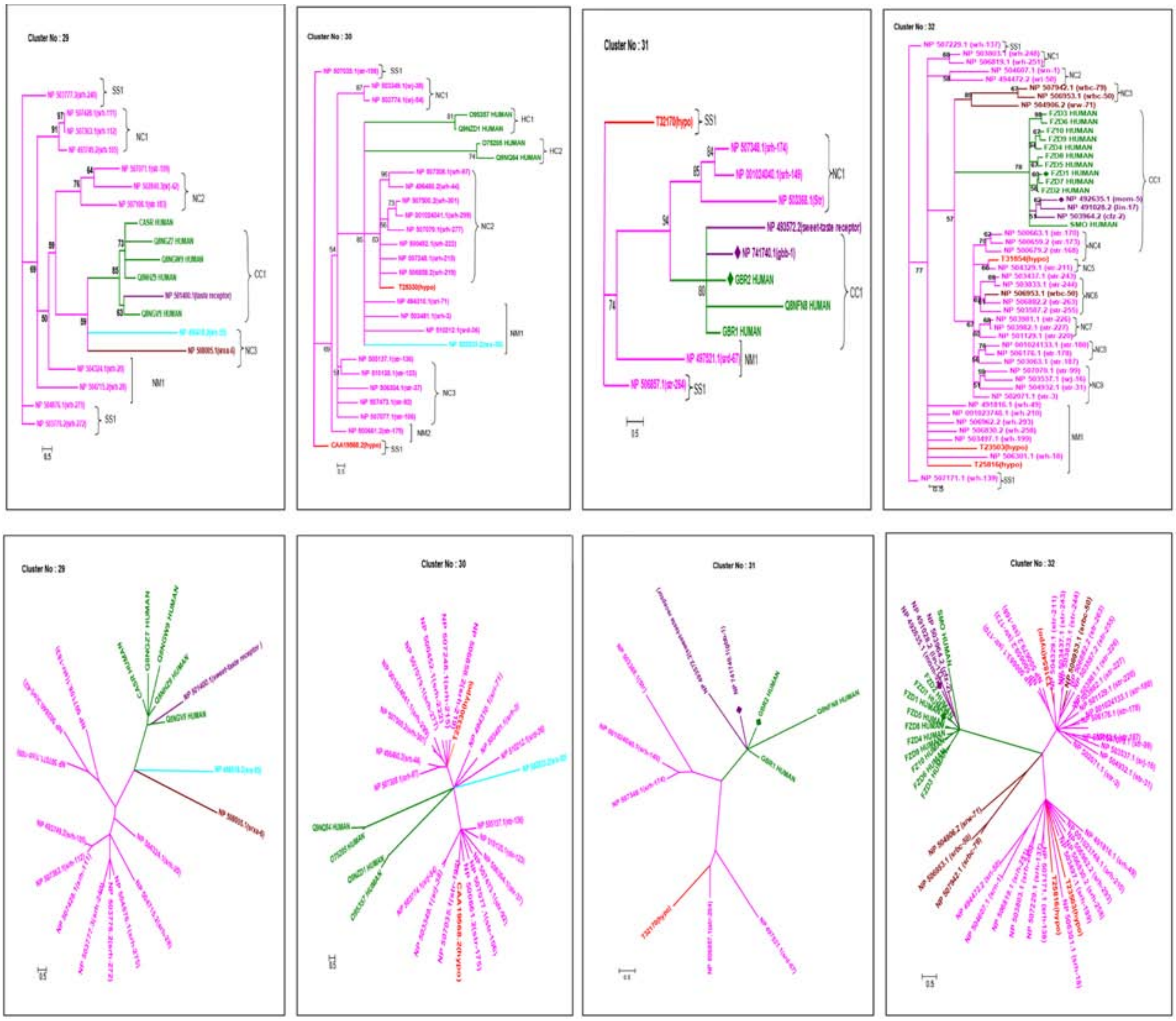

Figure S8. Same as Figure S1 but for glutamate receptors (Clusters 29–31), frizzled and smoothened receptors (Cluster 32).

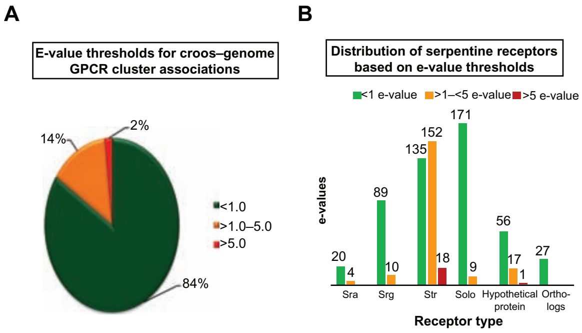

Figure S9. (**A**) Graphical pi-chart representation for the association of *C. elegans* GPCRs with 32 human GPCR profiles at different *E*-value thresholds: threshold of <1.0 for 84% (green color), >1 to <5 for 14% (orange color) and >5 for 2% (red color) in dataset. (**B**) Bar diagram illustrates the distribution of receptors at superfamily level, hypothetical protein, orthologs at different *E*-value thresholds: such as <1.0 (green color), >1 to <5 (orange color) and >5 (red color) in dataset.

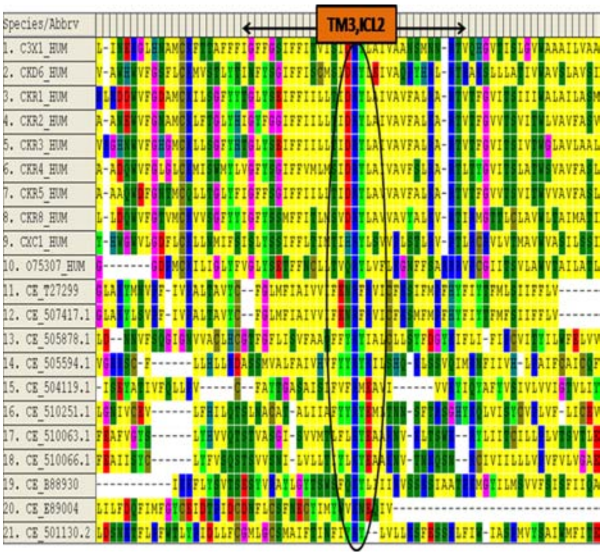

Figure S10. A snapshot of MSA for cluster 12. Alignment window of cluster 12 denotes the conserved E/DRY motif in human GPCRs whereas YRY motif in *C.elegans* GPCRs at the topology of TM3, ICL2.

Supplementary Tables

Table S1. List of Superfamilies and families for *C. elegans* Serpentine receptors (SR).

Table S2. List of predicted transmembrane helices by HMMTOP and SOSUI.

Table S3. List for Cluster no, Protein Id, associated E-value for correct association and false association from the trail study of known association with *Drosophila* GPCRs.

Table S4. Cluster-wise distribution of *C. elegans* GPCRs with respective Protein Id, Gene Id, Common name and associated E-value.

Table S5. List of orthologous GPCRs observed in human—*C. elegans* cross genome GPCR clusters.

Table S6. Cluster wise distribution for hypothetical proteins observed in human—*C. elegans* cross genome GPCR clusters.

Table S7. List of identified conserved amino acid residues in SR superfamilies.