

Research article

Open Access

## The repertoire of G protein-coupled receptors in the sea squirt *Ciona intestinalis*

N Kamesh, Gopala K Aradhyam and Narayanan Manoj\*

Address: Department of Biotechnology, Bhupat and Jyothi Mehta School of Biosciences Building, Indian Institute of Technology Madras, Chennai 600036, India

Email: N Kamesh - kamesh@smail.iitm.ac.in; Gopala K Aradhyam - agk@iitm.ac.in; Narayanan Manoj\* - nmanoj@iitm.ac.in

\* Corresponding author

Published: 1 May 2008

Received: 8 December 2007

BMC Evolutionary Biology 2008, 8:129 doi:10.1186/1471-2148-8-129

Accepted: 1 May 2008

This article is available from: <http://www.biomedcentral.com/1471-2148/8/129>

© 2008 Kamesh et al; licensee BioMed Central Ltd.

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

### Abstract

**Background:** G protein-coupled receptors (GPCRs) constitute a large family of integral transmembrane receptor proteins that play a central role in signal transduction in eukaryotes. The genome of the protochordate *Ciona intestinalis* has a compact size with an ancestral complement of many diversified gene families of vertebrates and is a good model system for studying protochordate to vertebrate diversification. An analysis of the *Ciona* repertoire of GPCRs from a comparative genomic perspective provides insight into the evolutionary origins of the GPCR signalling system in vertebrates.

**Results:** We have identified 169 gene products in the *Ciona* genome that code for putative GPCRs. Phylogenetic analyses reveal that *Ciona* GPCRs have homologous representatives from the five major GRAFS (*Glutamate, Rhodopsin, Adhesion, Frizzled* and *Secretin*) families concomitant with other vertebrate GPCR repertoires. Nearly 39% of *Ciona* GPCRs have unambiguous orthologs of vertebrate GPCR families, as defined for the human, mouse, puffer fish and chicken genomes. The *Rhodopsin* family accounts for ~68% of the *Ciona* GPCR repertoire wherein the LGR-like subfamily exhibits a lineage specific gene expansion of a group of receptors that possess a novel domain organisation hitherto unobserved in metazoan genomes.

**Conclusion:** Comparison of GPCRs in *Ciona* to that in human reveals a high level of orthology of a protochordate repertoire with that of vertebrate GPCRs. Our studies suggest that the ascidians contain the basic ancestral complement of vertebrate GPCR genes. This is evident at the subfamily level comparisons since *Ciona* GPCR sequences are significantly analogous to vertebrate GPCR subfamilies even while exhibiting *Ciona* specific genes. Our analysis provides a framework to perform future experimental and comparative studies to understand the roles of the ancestral chordate versions of GPCRs that predated the divergence of the urochordates and the vertebrates.

### Background

On a taxonomic and phylogenetic scale, *Ciona intestinalis* is a protochordate belonging to the ascidian class of urochordates that diverged from a lineage leading to the vertebrates approximately 520 million years ago [1]. This

extant ascidian occupies a crucial place in the "Tree of life" as an out-group to the vertebrates and hence studies addressing evolutionary aspects of *Ciona* have the potential to offer insight into some of the most intriguing questions about the origin of the vertebrates from a chordate

lineage. Recent genomic analysis has shown that the urochordates, and not cephalochordates, are the closest extant relatives of vertebrates [2]. As an added advantage, this ascidian is also amenable to experimentation and even has a century old history of being used for embryological studies [3]. A translucent morphology, availability of developmental mutants, established transgenic experimental procedures, EST databases and quickly spawning embryos are just a few advantages that make *Ciona* a favourite model system for developmental biologists. *Ciona* has a compact genome size of about 160 million base pairs and contains approximately 16,000 protein-coding genes [4].

GPCR based signal transduction is ubiquitous in eukaryotic genomes and forms the basis of detection of diverse environmental cues such as odorant molecules, amines, peptides, lipids, nucleotides and photons. A common structural feature of GPCRs is the presence of a highly conserved architecture of seven stretches of transmembrane spanning residues linked by alternate extracellular and intracellular loops. The diversity among GPCRs primarily stem from the presence of characteristic N-terminal extracellular domains and C-terminal intracellular domains and to a relatively lesser extent from the connecting loops which share limited sequence similarity. A number of fully sequenced genomes have been mined for their repertoire of GPCRs and comparative phylogenetic studies described for *Homo sapiens* [5], *Tetraodon nigroviridis* [6], *Anopheles gambiae* [7], *Gallus gallus* [8], *Rattus rattus* [9], *Mus musculus* [10], *Drosophila melanogaster* [11] and *Branchiostoma floridae* [12].

On the basis of a large scale phylogenetic analysis, GPCRs in human and subsequently in many other genomes have been classified into five major families (GRAFS); *Glutamate* (G), *Rhodopsin* (R), *Adhesion* (A), *Frizzled/Smoothed* (F), and *Secretin* (S) [5,6,8-12]. We denote the family names in italic with an initial capital letter to avoid possible confusion with for example, the rhodopsin receptor. HMMs (Hidden Markov Models) derived from the GRAFS classification system, along with the HMMs of other non-GRAFS families of GPCRs have been used to tract putative GPCRs from 13 completely sequenced genomes (including *Ciona*) into specific families and subfamilies. The analysis revealed that GRAFS families can be found in all bilateral species suggesting that they arose before the split of nematodes from the chordate lineage [13]. A recent review details progress in mining the gene repertoire and expressed sequence tags (ESTs) for GPCRs in several completed genomes [14].

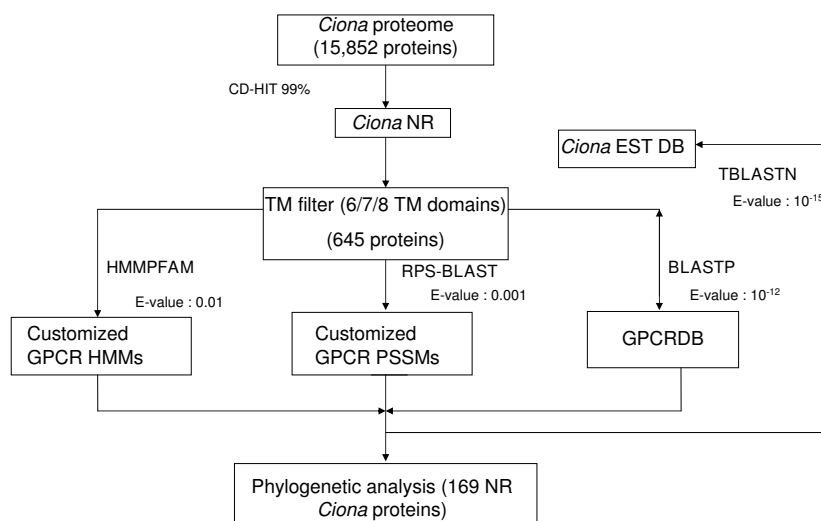
The *Ciona* genome provides scope for identification of GPCRs that are analogous in function to its vertebrate counterparts. This provides a resource for comparative

analysis aimed at assigning function to putative gene products [15]. Furthermore, comparative protein domain analysis as implemented in evolutionary trace methods can identify functional interface residues essential for ligand binding and subsequent signalling events [16]. Identification of conserved domains and/or novel GPCRs and their ligands are of major interest in light of the well-recognized roles of GPCRs in clinical medicine [17-19]. Furthermore, the repertoire of GPCRs in *Ciona* could provide insights into some of the intriguing questions in evolutionary biology about the origins and evolution of the GPCR signalling system in a protochordate and its further diversification into the vertebrate lineage. *Ciona* possesses organ systems are homologous to vertebrate heart, thyroid, notochord and pineal systems [20,21]. Our comparative genomic analysis could now serve as a basis for carrying out molecular genetic studies in *Ciona* to address many functional and regulatory aspects of GPCR associated vertebrate organ development and physiology. Finally, it is also important to note that such genome-wide comparisons may provide some wrong conclusions about orthologous relationships in individual cases and results from such analysis are best used as complementary tools for carrying out specific experimental studies.

Here, we describe the repertoire of GPCRs in the *Ciona intestinalis* genome and provide a comparative perspective to the GRAFS families, including a detailed analysis of the LDLRR-GPCR (Low-density lipo-protein receptor repeat containing GPCR) family in *Ciona*.

## Results

Our aim was to generate an independent data set of the repertoire of *Ciona* GPCRs and thereupon, using phylogenetic approaches provide a comparative genomic perspective of the salient features observed in the repertoire. In order to generate the complete set of putative GPCRs, we adopted a comprehensive strategy wherein predicted 6/7/8 transmembrane receptor sequences identified from the JGI (Joint Genome Institute, USA) *Ciona* proteome was subjected to an array of similarity and pattern search analysis for GPCR specific features. The hits from these search methods were further subjected to phylogenetic analyses (Figure 1). Such a process identified 169 putative GPCRs that represent ~1.1 % of the total number of gene products predicted from the *Ciona* genome. The proportion of GPCRs to the genome size is therefore comparable to those predicted in vertebrate and insect genomes [13]. To the best of our knowledge, manual searches revealed only 30 unique *Ciona* GPCR sequences to be present in the current updated public sequence repositories. The *Ciona* GPCRs identified are summarized in Table 1 and Table 2. The complete list of *Ciona* GPCR sequences identified through this study is available as Additional data file 1.



TM filter: includes HMMTOP, TMHMM and SOSUI set of Transmembrane prediction tools

### Figure 1

**GPCR sequence analysis strategy employed.** The *Ciona* Proteome was searched for 6/7/8 'TM' segment spanning sequences and the hits were taken for comparison against GPCRDB and against customized HMMs and PSSMs of GPCR families/subfamilies using an array of similarity/pattern search tools like BLASTP, HMMPFAM and RPS-BLAST. Subsequently, a number of phylogenetic analyses were performed on the sequences identified as putative GPCRs.

HMM based genome-wide GPCR surveys have been previously carried out in 13 genomes, including that of *Ciona* [13]. While this HMM based study provided a broad overview of the repertoire of GPCRs in the *Ciona* genome, an independent large-scale molecular phylogenetic treatment with respect to the tunicate GPCRs has not emerged and awaits investigation. GPCR features reflecting highly specialized ascidian specific biology are usually evident from such phylogenetic analysis. Of particular interest are whole-genome orthology and paralogy comparisons that are not evident from the previous HMM based search approach [13]. A table describing orthology and paralogy observations identified in *Ciona* GPCRs through our comparative genomic studies is provided [Additional data file 2: sheet 1]. Our phylogenetic study also provides the first ever insight into *Ciona* specific protochordate GPCR repertoire apart from vindicating the presence of the five major GRAFS families (Figure 2). Homologs of the chemosensory GPCRs from nematodes, plant GPCRs, Yeast Pheromone receptors, gustatory and olfactory receptors from insects were not identified. 66 clear orthologs of human GPCRs could be identified in *Ciona* based on phylogenetic analysis.

Since EST evidence covers three quarters of the predicted *Ciona* genes, we also identified EST matches for the 169 putative GPCRs using TBLASTN with a stringent cut off E-

value of  $1e^{-15}$ . Except for 12 of the 169 sequences, the rest of the GPCRs had at least one EST match. EST matches for these receptors were sorted as being derived from different developmental stages including fertilized egg, tailbud embryo, larval, embryo, gastrula and neurula, cleavage/cleaving stage, Stage 3, directional larva, young adult, juvenile and mature adult stage as well as adult tissues like heart, neural complex, gonad, digestive, blood/haemocytes, endostyle and testis [Additional data file 2: sheet 2]. A preliminary observation from the EST table also showed that as many as 133 receptors had at least one match that could be recovered in the ten different developmental stages of *Ciona* at the cut off E-value of  $10^{-15}$ , suggestive of a role for these GPCRs in developmental processes.

### The Rhodopsin Family (116 members)

The *Rhodopsin* family in *Ciona* includes 116 receptors that constitute about 68% of the entire GPCR repertoire. The *Rhodopsin* family has analogous subfamily level representatives among the four main groups, termed  $\alpha$ -,  $\beta$ -,  $\gamma$ -, and  $\delta$ -group, which in turn are subdivided into 13 major subfamilies in humans [5]. At the individual gene level, 39 out of the 116 *Rhodopsin* receptors display a clear human ortholog. However, unlike in the human *Rhodopsin* family, distinct clustering into the four major groups was not observed in *Ciona*. This is only expected since it is known

**Table 1: The number of G protein-coupled receptors predicted in the non-Rhodopsin GPCR families/subfamilies of *Ciona* in comparison with the human genome**

Receptor	Human *	<i>Ciona</i>
<b>Adhesion</b>		
CELSR	3	1
HE6/GPR126	1+1	1
GPR112/GPR128/GPR97/GPR56/GPR114	1+1+1+1+1	-
LEC/ETL	3+1	3
CD97/EMR/GPR127	1+3+1	-
BAI	3	-
GPR133/GPR144	1+1	-
VLGR1	1	-
GPI23/124/125-like	1+1+1	1
GPR110/111/113/115/116	1+1+1+1+1	-
Unclassified <i>Ciona</i> ADH	-	24
<b>Frizzled</b>		
Frizzled	10	3
Smoothened	1	1
TAS2	25	-
<b>Glutamate</b>		
CASR	1	1
GABA <sub>B</sub>	2	3
Metabotropic Glutamate	8	3
RAIG	4	1
Taste I	4	-
<b>Secretin</b>		
CALCRL/CRHR	2+2	1+2
GLPR/GCCGR/GIPR	2+1+1	3+0+0
PTHR	2	2
GHRHR/PACAP/SCTR/VIPR	1+2+1+1	-
cAMP GPCR-like	-	1
Methuselah-like	-	1
GPR107 (LUTR)-like	1	1

\* Numbers and abbreviations are as described in [5, 8].

that such grouping or relatedness is very restricted as regards species content.

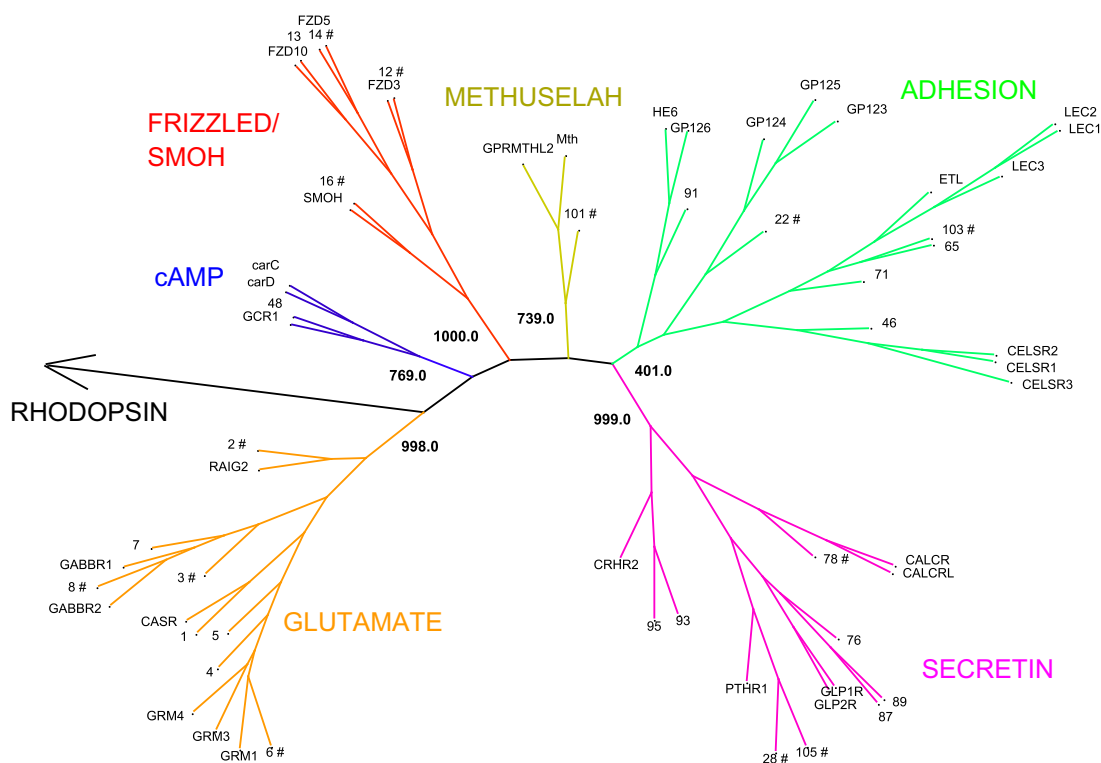
Our large-scale phylogenetic analysis of *Ciona Rhodopsin* family with annotated receptors from other genomes revealed the presence of 9 of the 13 subfamilies described for the human *Rhodopsin* family (Figure 3). One noteworthy finding in the *Ciona Rhodopsin* repertoire is the identification of an innovative low-density lipoprotein receptor repeat containing GPCR cluster (LDLRR-GPCR). The LDLRR-GPCRs, the *Ciona* LGR-like sequences (Leucine rich repeat containing GPCRs) and the GLHRs (Glycoprotein hormone receptor) clustered into one main branch with a high bootstrap support (999 out of 1000 NJ trees) (Figure 3). For the sake of clarity in discussions, the *Rhodopsin* phylogenetic representation is split into two sub-

**Table 2: The number of G protein-coupled receptors predicted in the Rhodopsin families/subfamilies of *Ciona* in comparison with the human genome**

Receptor	Human*	<i>Ciona</i>
<b>Rhodopsin (α)</b>		
Prostaglandin	15	2
Amine	40	8
Opsins I	9	3
Melatonin	3	3
MECA	22	5
Unclassified amine	-	6
Unclassified Opsin-like	-	3
Unclassified EDG/CNR-like	-	1
TRE1/GPR84	1	3
<b>Rhodopsin (β)</b>		
Peptide	43	9
Unclassified peptide	-	1
<b>Rhodopsin (γ)</b>		
SOG	15	1+0+2
MCHR	2	-
Chemokine cluster	42	2
Unclassified Chemokine receptor cluster-like	-	12
<b>Rhodopsin (δ)</b>		
MAS	8	-
LGR-like/GLHR	8	2+1
LDLRR	-	15
Unclassified LGR-like	-	14
Purine	42	-
Olfactory	388	-
Other Rhodopsins	23	23

\* Numbers and abbreviations are as described in [5, 8].

sets, one comprising the LDLRR-GPCR/LGR-like set of receptors and the other consisting of the non-(LDLRR-GPCR/LGR-like) cluster. Among the non-(LDLRR-GPCR/LGR-like) set of sequences, the melanocortin receptors of the α-group, the classical chemokine receptors (CCRs, CXCRs), melanocyte-concentrating hormone (MCHR) subfamilies of the γ-group and the purine (PUR), MAS-related receptor (MRG) and olfactory subfamilies of the δ-group are conspicuously absent. As many as 23 receptors could not be included in any subfamily with significant phylogenetic support. These sequences were thus designated as "Other Rhodopsins". *Rhodopsin* subfamily members that could not be annotated at the individual gene level due to lack of support *via* cross-genome phylogenetic clustering were deemed as "Unclassified Rhodopsins". The "Unclassified Rhodopsins" are different from the "Other Rhodopsins" in that most of these "Unclassified Rhodopsins" could be included into individual *Rhodopsin* subfamilies when analyzed with limited number of sequences, but in a larger data set do not give stable or consistent topology.



**Figure 2**  
**Phylogenetic relationship between GPCRs in *Ciona* and other genomes.** The figure illustrates the presence of representatives of *Glutamate*, *Rhodopsin*, *Adhesion*, *Frizzled*, and *Secretin* family members in *Ciona*. Two sequences that are homologous to cAMP and Methuselah GPCRs are also represented. The position of the *Rhodopsin* family was established by including 15 receptors from the *Rhodopsin* family. The divergent, "Other/Unclassified" GPCRs known to lack reliable homologs from other species or which are fast evolving, are excluded from the final representation. For display reasons bootstrap values are not represented in the figure. Instead, the corresponding tree file in standard Newick format is provided [Additional data file 5]. For Figure 2, and Figures 3, 5 and 6, the multiple sequence alignment was built taking into account terminally truncated TM spanning regions, while the consensus phylogenetic tree was calculated using NJ method on 1000 replicas of the dataset. *Ciona* GPCR taxons are represented in numerals as per numbering in Additional data file 2. Abbreviations for known GPCRs are as described in [5, 6, 8, 9] and based on Swiss-prot IDs. *Ciona* GPCRs that deviate from the predicted 7TM structure are marked using a '#' symbol.

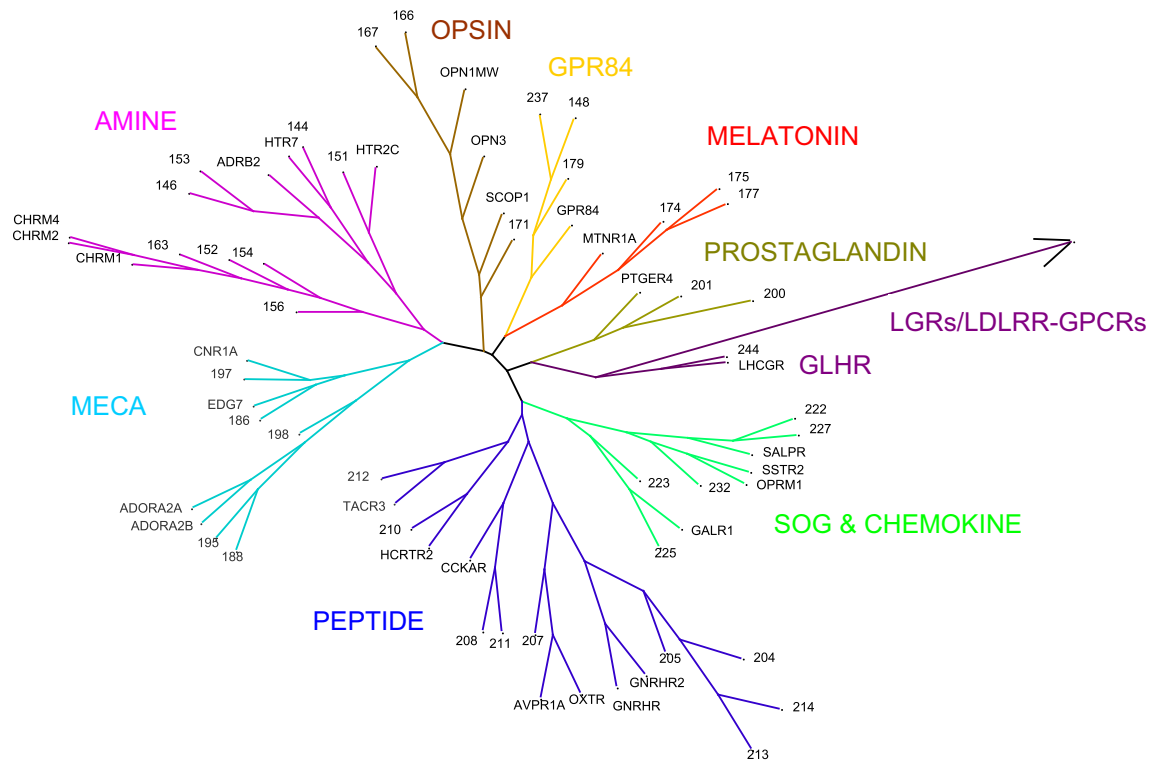
Since several of these "Unclassified *Rhodopsins*" and "Other *Rhodopsins*" appear to be *Ciona* specific and were divergent from the *Rhodopsin* subfamilies in other genomes, they were excluded from the final phylogenetic representation to avoid tree artefacts.

**The LDLRR/LGR-like/GLHR cluster (32 members)**

*Ciona* shows a novel lineage specific expansion and innovation in the LDLRR-GPCR/LGR-like/GLHR *Rhodopsin* cluster with 32 receptors, 30 of which are similar to the human INSL3/relaxin binding GPCRs in their TM regions. Of the 32 receptors in this cluster, two sequences were found to be LGR-like (ci0100148288 and ci0100151424), with Leucine-rich repeat domains (LRR) observable in their N-termini. The sequence ci0100148288 possesses five LRR domains and was iden-

tified as a candidate INSL3/Relaxin GPCR-like based on conserved domain database searches. The sequence ci0100151424 is an N-terminal LRR containing GPCR that does not bear any significant similarity to INSL3/Relaxin binding GPCRs in its TM region. One glycoprotein hormone receptor could be recovered (ci0100133821) and was identified as an ortholog of LHCGR (Lutenizing hormone/chorio-gonadotropin receptor).

It is noteworthy that as many as 14 out of the 32 receptors apart from bearing sequence similarity to INSL3/relaxin binding GPCRs in their TM region, comprise of a low density lipoprotein receptor Class A domain (LDL-A) represented either singly or in 2–10 multiple tandem repeats at the N-termini [Additional data file 3]. In contrast to other

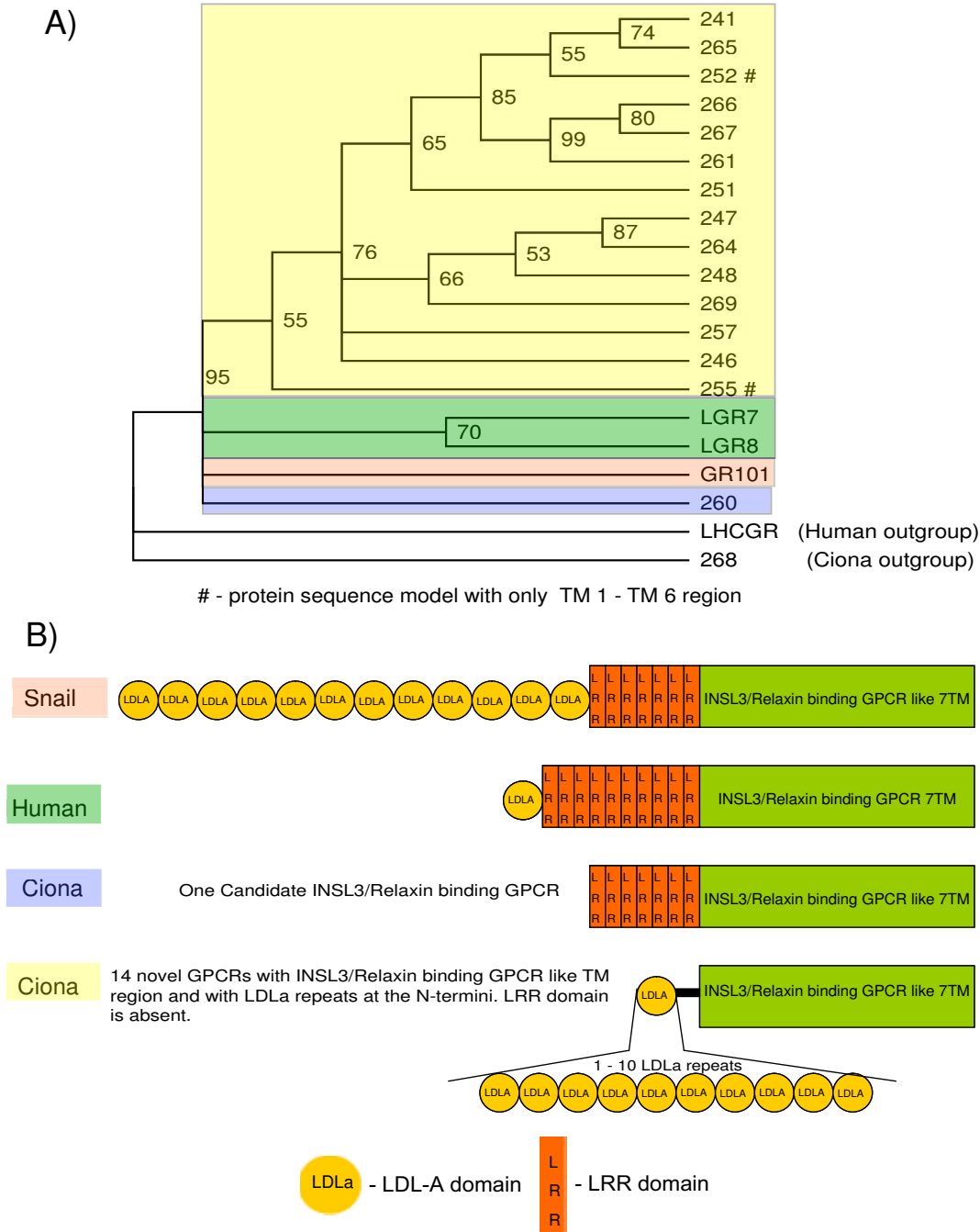


**Figure 3**  
**Phylogenetic relationship between non-(LDLRR-GPCR/LGR) Rhodopsin receptors in *Ciona* and other genomes.** All members of the *Ciona* Rhodopsin subfamilies except for the LDLRR-GPCRs/LGRs were included with closely related sequences from other genomes to construct the phylogenetic tree. To ascertain the phylogenetic position of *Ciona* LGR-like/LDLRR-GPCR cluster, 10 of those sequence members were added for tree reconstruction and later the branches removed from the final representation to be replaced by an arrow. The divergent "Other/Unclassified Rhodopsins" known to lack reliable homologs from other species, or which are fast evolving were excluded from the final representation. For display reasons boot strap values have not been shown. Instead, the corresponding tree file in standard Newick format is attached [Additional data file 6]. *Ciona* GPCRs that deviate from the predicted 7TM structure are marked using a '#' symbol.

vertebrate LGRs (LGR7/LGR8) where a single LDL-A module at the N terminus of the ectodomain is followed by multiple LRRs and 7 TM helices, the 14 *Ciona* LDLRR-GPCR family members lack LRRs [22] [Additional data file 3]. Also notable is the fact that this unique domain organisation of LDL-A domain tandem repeats in *Ciona* has been reported only in snail LGR that contains tandem repeats of LDL-A domains in addition to multiple LRR domains [23] [Swiss-Prot: GR101\_LYMST]. 14 additional sequences belonging to this subfamily possess shorter N-terminal regions that do not contain any explicitly recognizable domains while bearing INSL3/Relaxin binding GPCR-like TM regions [Additional data file 3]. These sequences are therefore not authentic N-terminal associated "Leucine Rich Repeat containing GPCRs" and hence been designated as "Unclassified LGR-like". It is possible that the 14 "Unclassified LGR-like" sequences are fragmentary or incompletely modelled at the N-termini. At

least one instance of an incorrectly aligned domain architecture in a LDL-A domain bearing GPCR was revealed (ci0100131758) after a BLAT of the sequence was performed against the *Ciona* genome. A proposed correct model is provided in the Additional data file 1.

A phylogenetic analysis was carried out to resolve the relationship between the human LGRs (LGR 7/LGR 8), the snail LGR, the 14 *Ciona* LDLRR-GPCR family members (excluding an incorrectly aligned member: ci0100131758) and the one candidate INSL3/Relaxin binding GPCR-like sequence (ci0100148288). The unifying theme among all these receptors is their INSL3/Relaxin binding GPCR-like TM region. This phylogenetic tree was rooted using a human LHCGR and a *Ciona* orphan LGR-like (ci0100151424) that are distantly related to these sequences [Figure 4]. The receptors in this data set were truncated to include only the sequence of the



**Figure 4**  
**Phylogenetic relationship of LDLRR-GPCR/LGR-like members of *Ciona* with those of Snail and Human LGRs and their domain organization.** A) Maximum likelihood tree of the TM regions of the 14 *Ciona* LDLRR-GPCRs, the *Ciona* INSL3/relaxin receptor-like and the Snail and Human LGRs. The tree is rooted using a human LHCGR and a *Ciona* orphan LGR (ci0100151424) that are distantly related to these sequences. Support values are indicated in percentages. Taxons represented in numerals refer to *Ciona* GPCRs as per information in Additional data file 2. *Ciona* GPCRs that deviate from the predicted 7TM structure are marked using a '#' symbol. B) Schematic diagram representing the modular domain organization corresponding to the clusters identified in the phylogenetic tree.

intervening TM 1- TM 6/7 region. This truncation was carried out to eliminate the variabilities in the number of LDL-A and LRR tandem repeats.

Phylogenetic analysis shows that the human LGRs, the snail LGR, the 14 *Ciona* LDLRR-GPCRs and the one *Ciona* INSL3/Relaxin GPCR-like (ci0100148288) clustered into one group separately from the distantly related human LHCGR and the *Ciona* LGR-like outgroup suggestive of a common evolutionary origin to the INSL3/Relaxin GPCR-like 7 TM region. Furthermore, the domain analysis also suggests that the 14 *Ciona* LDLRR-GPCRs and the one candidate *Ciona* INSL3/Relaxin GPCR-like seem to have evolved independently of the human LGRs and the snail LGR in the non-TM N-terminal regions.

#### **The Glutamate family (8 members)**

Many *Glutamate* (GLR) receptors are crucial modulators involved in neurotransmission. Most of these receptors have long N-termini that serve to bind a cognate ligand [24]. The *Ciona* *Glutamate* family clustered with a strong bootstrap support (998/1000) separating the cluster from its closest neighbours (Figure 2). Phylogenetic analysis revealed that the *Ciona* genome contains 8 members of the *Glutamate* family. These receptors represent all of the GLR subfamilies with the exception of Taste (TAS1) receptors. Human orthologs for Calcium-sensing receptor (CASR), Retinoic acid inducible GPCR (RAIG), Metabotropic glutamate receptor (GRM) and GABA receptors could be identified in *Ciona* with significant sequence similarity. Among the *Ciona* GLRs, the CASR (ci0100130340) homolog displayed the highest identity to the human counterpart CASR (34.7%). ESTs for the metabotropic GABA<sub>B</sub> related receptor (ci0100152670) could be retrieved from the neural-complex tissues in *Ciona*, suggesting similar roles for this receptor in synaptic function as proposed in vertebrates [Additional data file 2: sheet 2].

#### **The Secretin family (8 members)**

The *Secretin* family includes receptors for calcitonin (CALCR), vasoactive intestinal peptide (VIPR2), glucagon-like peptide (GLPR), pituitary adenylate cyclase-activating polypeptide (PACAP), the parathyroid hormone receptor (PTHr) and several other related peptides as well as hormones [5]. The *Ciona* *Secretin* family received very high bootstrap support (999 out of 1000 NJ trees) (Figure 2). 8 GPCRs could be identified in *Ciona* and are related to the vertebrate *Secretin* family with which they shared high sequence similarity. The phylogenetic clustering also supports possible specific gene duplication events for three pairs of sequences, evident from the high bootstrap support (>80%) for the tree nodes for the pairs (ci0100139945/ci0100151327), (ci0100145252/ci0100145281), and (ci0100145584/ci0100145837).

This is consistent with the recent report on the evolution of *Secretin* family receptors in *Ciona* [25]. The gene pair (ci0100139945/ci0100151327) clusters with the PTHR subfamily of receptors, while the pair (ci0100145252/ci0100145281) clusters with GLPR group. The remaining pair (ci0100145584/ci0100145837) clusters with the CRHR group, whereas ci0100141310 and ci0100141557 appear to be orthologs of GLPR and CALCR/CALCRL respectively (Figure 2). Among *Secretin* receptors, the *Ciona* PTHR homologs (ci0100139945/ci0100151327) showed the highest sequence identity (37% identity to PTHR1 in both cases) to their human counterparts (Additional file 4).

#### **The Frizzled/Smoothened family (4 members)**

The *Ciona* proteome data set contains 4 members of the *Frizzled/Smoothened* GPCR family each of which have at least 6 TM regions. The *Ciona* *Frizzled* receptor cluster branched out with a very high bootstrap support (100%) in NJ trees (Figure 2). *Frizzled* receptors have been shown to couple with G-proteins after activation by Wnt, a glycoprotein ligand and are known to play a key role in tissue-polarity and cell signalling [26,27]. The *Frizzled* family is a recently identified group among the GPCRs, although its origin can be traced back to the arthropods. *Ciona* possesses a clear one-to-one ortholog each for human FZD3, FZD5 and FZD10 apart from a smoothened-like (SMOH) ortholog. The conserved (K-T-X-X-X-W) motif required for Wnt/ $\beta$ -Catenin signalling pathway can be observed two amino acids after the seventh transmembrane helix in all the recovered *Ciona* *Frizzled* receptors [28]. An alignment of the TM regions of the *Frizzled* ortholog pairs (FZD) of humans and *Ciona* revealed that they are very well conserved with identities in range of ~48–60 %. The SMOH receptor shares ~57 % identity with its *Ciona* ortholog (ci0100150930) in the TM region. However, the Taste2 (TAS2) receptors seem to be absent in *Ciona*.

In a separate analysis, we identified four more *Ciona* *Frizzled* family members (ci0100152761, ci0100150136, ci0100138855 and ci0100153324) that were fragmentary models with less than 4 TM regions. These fragmentary sequences were not included in the final *Ciona* GPCR dataset for better handling of the MSA (Multiple sequence alignment) data for subsequent phylogenetic studies.

#### **The Adhesion family (30 members)**

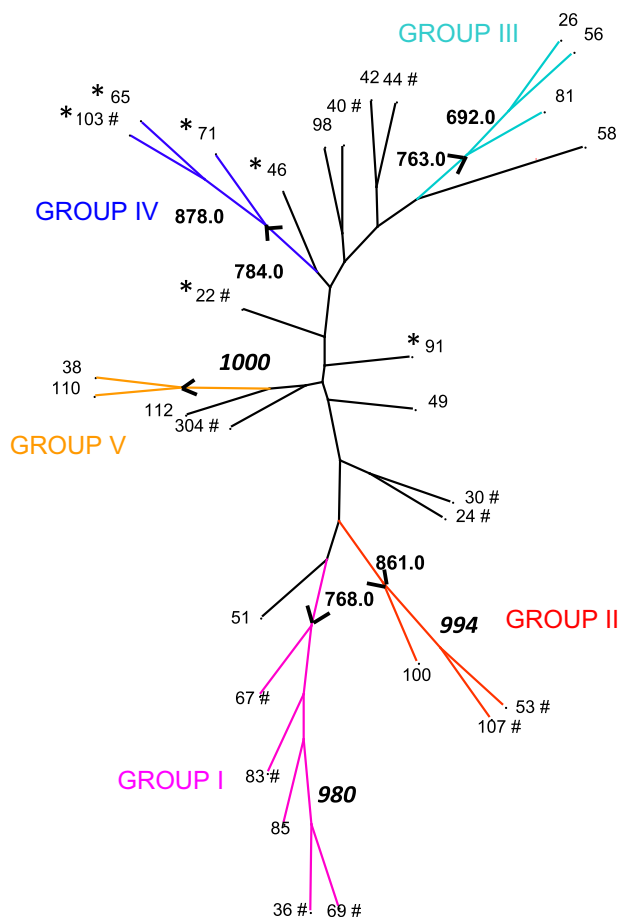
The *Ciona* genome includes 30 *Adhesion* family members and is the second largest group of GPCRs after the *Rhodopsin* family. This family is characterized by long N-termini that contain multiple functional domains with adhesion-like motifs that are often sites of glycosylation events [29]. It is noteworthy that *Ciona* with a relatively smaller genome possesses as many *Adhesion* members as any of the other larger vertebrate genomes [5,6,8]. The



density (number of *Adhesion* genes/genome size) of GPCR *Adhesion* family is 1 gene/~5 Kb in *Ciona* whereas it is 1 gene/~90 Kb for human. This denser *Ciona* GPCR group relative to human genome is also accompanied by a lack of clear-cut homology to human *Adhesion* GPCRs. Among all the major family clusters in our analysis, the *Adhesion* family received the lowest bootstrap support (401 out of 1000 NJ trees) (Figure 2). Of the 30 *Ciona* receptors identified, 6 showed unambiguous orthologous relationship with human members in a large scale comparison of the *Adhesion* GPCRs of the two genomes (data not shown). *Ciona Adhesion* members that received complete support from its human counterparts were included in the final representation (Figure 2). The human genome contains 33 *Adhesion* receptors that display a phylogenetic clustering into eight main groups (I-VIII) [29]. *Ciona* orthologs of the human *Adhesion* subfamilies includes one representative (ci0100145494) of the HE6/GP126 cluster (human group VIII), one representative (ci0100130008) of the GPR123-125 cluster (human group III) and one ortholog (ci0100132112) of the CELSR genes (human group IV) [29]. The remaining 3 receptors (ci0100137028, ci0100140016 and ci0100150579) are orthologous to the LEC subfamily of human *Adhesion* group I.

A search for potential orthologs of the remaining 24 *Ciona* receptors using PIPEALIGN and further phylogenetic analysis with top hits from the search failed to identify unambiguous orthologs when performing both NJ and Maximum Likelihood analysis. *Adhesion* GPCRs that could not be assigned specific subfamilies due to lack of support in cross-genome phylogenetic analysis were deemed as "Unclassified *Adhesion*". 21 of the 24 *Ciona* "Unclassified *Adhesion*" models had only the GPS proteolytic domain in the N-termini. Among the remaining three "Unclassified *Adhesion*" receptors, ci0100130017 had a LRR domain while ci0100132869 and ci0100131580 protein models possess EGF repeats in the N-termini.

A phylogenetic analysis of the *Ciona Adhesion* GPCRs was performed on the region from the TM 1 to TM 7 of the receptors. The analysis revealed that the *Ciona Adhesion* family clustered into five different groups each receiving more than 75 % bootstrap support in (1000 bootstraps) NJ trees (Figure 5). The five clusters were denoted as Group I to Group V. This group assignment of the *Adhesion* GPCRs is *Ciona* specific and is independent of the human *Adhesion* groups referred to by Bjarnadottir et.al., 2004 [29]. 16 *Ciona Adhesion* receptors could be included into these five groups reflecting a probable paralogous relationship between the groups. The remaining *Adhesion* receptors remain as outgroups and do not cluster with sufficient phylogenetic support with other *Ciona* members. The phylogenetic analysis indicates that *Ciona Adhesion*



**Figure 5**  
**Phylogenetic relationship within the *Ciona* GPCR set of *Adhesion* receptors.** A phylogeny of *Ciona Adhesion* GPCRs shows the presence of five different paralogous clusters (group I- group V). Tree branches that support the five groups (> 75% bootstrap support) are coloured and their bootstrap values are represented. Branches that suggest gene duplication events (>98% bootstrap support) are reported with the bootstrap values in bold italics. *Ciona Adhesion* GPCR homologs of Human members are depicted with a '\*' symbol and the GPCRs that deviate from the predicted 7TM structure are marked using a '#' symbol.

repertoire constitutes a very diverse set of receptors sharing relatively poor sequence identity (Figure 5). The three members in Group III share the least identity (20%) while the two members in Group V were the most identical (89%). The Group V receptors are presumably gene duplication events as evident from the high bootstrap support (100%) of that node. Three members belonging to Group II share 25 % identity, while the five members of the Group I and the three members in Group IV share 23 % and 25 % identities respectively. In all cases, only the

identities within the TM regions were considered. Group IV constitutes receptors that are orthologous to human LEC/ETL subfamily of receptors. In addition to the gene pair (ci0100152766/ci0100130804) in Group V, our analysis also lends support to a hypothesis of two more probable gene duplication events for the tree nodes (ci0100130776/ci0100139268) and (ci0100134612/ci0100152008) (Figure 5). Sequence alignment of the TM regions revealed that the ci0100130776/ci0100139268 pair receptors were 57% identical while the ci0100134612/ci0100152008 pair shares an identity of 52 %.

### Specialized tunicate GPCRs

Our analysis revealed that the *Ciona* genome contains at least two representatives that could not be classified as unambiguous orthologs of the GRAFS families. These include one receptor (ci0100132129) that is similar to the protist *Dictyostelium* cAMP/Crl GPCR family and while the other receptor (ci0100150391) is homologous to the insect related Methuselah protein. The two *Ciona* receptors cluster with the corresponding cAMP and Methuselah GPCRs in a cross-genome phylogenetic analysis with a strong bootstrap support of 769 and 739 respectively, (out of 1000 replicates) in NJ trees (Figure 2). cAMP receptors in *Dictyostelium discoideum* mediate the coordinated aggregation of individual cells into a multi-cellular organism and regulate the expression levels of a number of developmentally regulated genes [30-32] while Crl receptors are implicated in cell growth regulation and tip formation in developing aggregates [33]. The Methuselah GPCR in *Drosophila* plays a role in ageing and increased resistance to several forms of stress including heat, starvation and oxidative damage [34]. The cAMP/Crl receptors are believed to be represented exclusively in protists, plant and fungal kingdoms while the Methuselah GPCR is thought to be present only in insects.

A BLASTP search for closest homolog of the *Ciona* cAMP GPCR-like sequence in the non-redundant NCBI database revealed that the *Ciona* sequence is most identical to predicted proteins in sea anemone (*Nematostella vectensis*) [Swiss-Prot: [A7SN55](#)] and *Danio rerio* [GenBank: [XP\\_001332705](#)] with an overall sequence identity of 42% for both cases. The recovered *Nematostella* sequence was observed to have only 5 TM regions. An alignment of the *Ciona* receptor and its sea anemone homolog showed an identity of nearly 37% when only the TM regions were considered. Among the *Dictyostelium* cAMP/Crl family members, the *Ciona* sequence ci0100132129 shows best identity with CARC (~21% identical in the TM regions), while the best identity among the Crl members is CrlA (~19 % identical in TM regions). Likewise, a BLASTP search with the *Ciona* Methuselah-like homolog (ci0100150391) with other insect genomes revealed that

it is ~23% identical in the TM region to a MTH 10-like receptor in the *Apis mellifera* genome (PIR:UPI000DB6BC5). However, the ectodomain that characterizes the *Drosophila* Methuselah GPCR is absent in the *Ciona* homolog.

### Discussion

Our phylogenetic analysis includes a collection of 169 putative *Ciona* GPCR sequences and highlights in detail the remarkable level of orthology between the protochordate and vertebrate GPCRs. Our analysis clearly suggests that a majority of the identified *Ciona* GPCRs have a clear vertebrate and in particular, a human homolog at the level of the major GRAFS subfamilies. Our results also show that at the individual gene level there are as many as 66 clear orthologs of human GPCRs in *Ciona* (Additional file 5). Given that *Ciona* possesses ancestral GPCR versions of the higher vertebral counterparts, there are also remarkable instances of lineage specific evolution as seen in the *Rhodopsin* and *Adhesion* families. From a protochordate perspective, the major differences from the human GPCRs in the *Ciona* repertoire can be attributed to the presence of 23 "Other Rhodopsins", the *Ciona* specific gene innovations of the LDLRR-GPCR cluster (14 sequences) and the unusually large *Adhesion* family (24 *Unclassified Adhesion* sequences). The presence of many "Unclassified/Other" GPCRs for which reliable orthologs could not be identified (*via* phylogenetic analysis) in other species, indicate that these are most likely ascidian specific genes. The other possibility is that these genes are evolving at a rapid rate, so much so that reliable orthologs could not be detected. In our final analysis and depiction of phylogenetic trees, we chose not to include these sets of "Unclassified/Other" sequences to avoid artefacts that can arise out of including sequences that are either evolving at a rapid rate or for which intermediate homologs could not be identified. Instances of incomplete/poorly modelled genes can also be a cause for the failure to detect reliable homologs for many of these GPCRs. With further sequencing and analysis of GPCRs in other protochordate genomes, it is likely to become clear if these genes are indeed protochordate related diversifications.

Our analysis of the *Ciona* GPCR family suggests that while the number of *Adhesion* receptors are comparable to that in vertebrates, the other GPCR families/subfamilies excepting that of the LDLRR-GPCRs, are represented at nearly 2 to 5 fold fewer numbers compared to the humans [Table 1, Table 2]. This observation is only expected for a genome that is 20 times smaller than the human genome.

### The Rhodopsin Family

Among the *Rhodopsin* receptors, the human  $\alpha$ - and  $\beta$ -group subfamilies are analogously represented in *Ciona* while among the seven subfamilies of the human  $\gamma$ - and

$\delta$ -groups, orthologs of the PUR, classical CHEM, MCHR and MRG receptor subfamilies could not be identified by phylogeny (Additional file 6). Simple BLAST searches identified many *Ciona* receptors as being similar to chemokine receptor cluster, but they did not receive sufficient support in the phylogenetic analysis. These were annotated as "Unclassified Chemokine receptor cluster-like". It is possible that these are highly divergent homologs of the chemokine receptor cluster genes. The absence of a clear classical chemokine (CCRs, CXCRs) receptor ortholog in *Ciona* agrees with previous finding that the evolution of chemokine receptors arose after the split of the vertebrates from the protochordate lineage [35].

*Ciona* has two representatives belonging to the prostaglandin receptor cluster. Although *Drosophila* has one remote homolog of a prostaglandin receptor, a readily recognizable homolog of a vertebrate-like prostaglandin is first observed only in the protochordate lineage. The sequence identities of *Ciona* prostaglandin receptors with human PTGER4 range from  $\sim 21 - 23\%$ . Prostaglandin receptors mediate a wide variety of actions and play important physiological roles in the cardiovascular and immune systems and in pain sensation in peripheral systems [36]. Reliable ESTs can be recovered from neural tissues for the PTGER homolog (ci0100153844) suggesting that these receptor homologs are involved in neural tissue physiology [Additional data file 2: sheet 2].

Among the peptide receptors, *Ciona* has a single copy of an authentic Tachykinin receptor which has been proposed to play an important role in feeding and sexual behaviour of the ascidians [37]. Furthermore, putative receptors for peptides like GnRH, Cionin (CCK/gastrin-related peptide), Oxytocin/Arginine-vasopressin and Hypocretin/Orexin were recovered in *Ciona* [Additional data file 2]. Among the *Ciona* GPCRs the GnRH receptor remains the best studied and is the only known example of an authentic protochordate gonadotropin receptor to date [38]. A GnRHR-like protein in *Drosophila* was later found to be activated by AKH, a peptide structurally similar to GnRH, thus making *Ciona* the only available protochordate model to possess authentic GnRHRs [39]. So far, three GnRHR-like *Ciona* homologs have been found to be active in assays and are known to stimulate both the cAMP and IP3 signalling pathways. A cannabinoid receptor homolog (ci0100149095) and a receptor (ci0100136887) that is most closely related to the lysophospholipid receptors were also identified. We could not recover cannabinoid receptor homologs from the genomes of protostomes, while homologs were found in echinoderms suggesting that these receptors have their origins in a common deuterostomal ancestor to echinoderms and chordates. ESTs for the EDG homolog (ci0100136887) were recovered from blood cells/haemo-

cytes suggesting that these receptors may have a role in the nervous system similar to that observed for vertebrate EDG receptors [40,41].

#### **Structural Innovations in the *Ciona* LDLRR/LGR-like GLHR Rhodopsin cluster**

*Ciona* shows a notable lineage specific gene innovation and gene expansion (14 confirmed members) of the LDLRR-GPCRs (Low density lipoprotein receptor repeat containing receptors). These receptors have TM regions that are homologous to vertebrate INSL3/Relaxin binding GPCRs. Lineage specific gene expansions have long been observed in several cases including that of the opsin and fish odorant receptor expansions in puffer fish [6] and trace amine receptor expansion in zebrafish [42]. The LDLRR-GPCR subfamily in *Ciona* displays an unusual combination of LDL-A domain tandem repeats in the N-termini. Although the presence of a single LDL-A domain in GPCRs was first observed in mammals, our unpublished study shows that the LDL-A domain repeats in GPCRs exist in protostomes (Molluscs, Arthropods) and in early deuterostomes like echinoderms (*Strongylocentrotus purpuratus*) and *Ciona* suggesting that the origin of these receptors can be traced back to a common bilateral ancestor. The presence of LDL-A repeats in mosaic proteins has long been known and is an example that supports the exon shuffling theory [43,44].

Based on phylogeny and domain architecture analysis we propose that while the signatures for the LDL-A domain and LRR domain containing *Rhodopsin* subfamily first arose in a bilateral metazoan ancestor, this subfamily has undergone several lineage specific innovations at the level of the combinations and tandem repeats of these domains. Our analysis suggests that the 14 *Ciona* LDLRR-GPCRs (in which the LRR domains are absent) have evolved divergently as a group [Figure 4, Additional data file 3]. This divergence is evident in the phylogenetic analysis of the TM regions of this group with those of the snail and human LGRs (Figure 4). It is possible that changes in the TM regions presumably accommodates for the changes that have occurred in the *Ciona* primary ligand binding N-terminal region that contain LDL-A domains but lack LRRs.

In vertebrate INSL3/relaxin binding GPCRs, a combination of a single cysteine-rich LDL-A domain and multiple LRRs at the ectodomain site is necessary for ligand binding, receptor activation and further downstream signalling from these receptors [45]. Receptors that lack the LDL-A domain were found to bind relaxin, but were unable to effect cAMP accumulation [46]. The LRR region in the ectodomain is essential for binding of glycoprotein hormones [47]. Notwithstanding the incomplete sequence information in the case of the 14 "Unclassified

LGR-like" sequences that do not have any identifiable domains at the N-termini, it is evident that the architecture of the 14 confirmed *Ciona* LDLRR-GPCRs that lack LRR domains is very different from vertebrate type C LGRs (LGR7/LGR8). These receptors are thus functionally unrelated to vertebrate INSL3/relaxin peptide binding receptors.

A typical cysteine-rich LDL-A domain found in human LDL receptor and human LDL Receptor related Protein (LRP) consists of ~40 amino acids containing six cysteine residues at fixed positions with the region between the fifth and sixth cysteines rich in aspartic and glutamic acid residues. These acidic residues in LDL receptors are known to interact with basic amino acids in lipoproteins [48,49]. The human LDL receptor contains at least 7 class A repeats [50,51] while the LRP contains clusters of 2, 8, 10 and 11 repeats [52]. Furthermore, in the LRPs, these repeats bind to large protein complexes like the  $\alpha_2$  macroglobulin-protease complexes and urokinase-type plasminogen activator-plasminogen activator inhibitor (uPA-PAI) [53,54]. A majority of LDL-A-like repeats in the 14 *Ciona* LDLRR-GPCRs conform to the classical LDL-A repeat structure. It is thus tempting to speculate that these domains are crucial for the specific function of the receptors and that these receptors likely bind to lipoproteins or large protein complexes as ligands, analogous to the classical human LDL receptor. This hypothesis based on comparative domain analysis may serve as a pointer towards further molecular studies. A similar hypothesis that the snail GRL101 receptor which contains both LDL-A and LRR repeats might be a member of a putative class of GPCRs that directly transduce signals carried by large extracellular (lipo)protein (complexes), has been proposed earlier [23]. The differences between the snail LGR-like and the *Ciona* LDLRR-GPCRs lie in the possibility that the snail receptor which contains LRRs can potentially offer a different ligand presentation mechanism compared to the *Ciona* receptors that lack LRRs.

Although direct evidence of lipoproteins as GPCR signalling agents is not available, previous evidence suggests that LDL molecules at physiological concentrations can activate phosphatidylinositol signalling and mobilization of calcium from intracellular stores in the absence of LDL receptor mediation [55,56]. Similarly, HDL stimulation of cAMP formation has also been demonstrated [57]. However, to the best of our knowledge, no authentic GPCR has been demonstrated to bind to lipoproteins as signalling molecules and therefore it should be a crucial step forward to de-orphanize this family of GPCRs in *Ciona*. A search of the *Ciona* EST database for matches to the novel *Ciona* LDLRR-GPCR protein models revealed that a majority of the hits are derived from the blood tissue and haemocytes [Additional data file 2: sheet 2].

Identification of three putative *Ciona* orthologs of vertebrate INSL-relaxin-like genes has recently been reported [58]. Furthermore, the authors hypothesize that the *Ciona* INSL/RLN orthologs most likely interact with a *Ciona* RTK-receptor and not a GPCR and the receptor switch from RTK-receptor to a GPCR occurred after the split of *Ciona* from the common lineage leading to the vertebrates [58]. From our own analysis of the 32 *Ciona* receptors in the LDLRR-GPCR/LGR-like/GLHR cluster, we found two LRR bearing LGR-like sequences (ci0100148288 & ci0100151424). Conserved domain searches of these two receptors further narrowed down only one candidate INSL3/relaxin receptor-like (ci0100148288) sequence. A phylogenetic analysis of *Ciona* sequence (ci0100148288) with vertebrate type C INSL3/Relaxin receptors (Figure 4) revealed that the *Ciona* receptor is not orthologous to vertebrate type C LGRs and is therefore unlikely to bind to vertebrate INSL3/Relaxin versions. Our analysis does not however exclude the possibility that the two LGR-like *Ciona* sequences (ci0100148288 & ci0100151424) can bind to the *Ciona* INSL-relaxin orthologs.

#### The *Ciona* GPCR Adhesome

Unlike vertebrates which possess multiple functional *Adhesion* N-terminal domains that include lectin, cadherin and laminin domains among others, the *Ciona* GPCR adhesome possesses a less complex domain organisation with a single GPS domain in 21 out of 30 receptors. Furthermore, these receptors did not correspond with any clear ortholog in other vertebrates. Orthologs for the vertebrate CELSR, LEC, GP123-GP125 and HE6 class of receptors could be observed but orthologs of other *Adhesion* members previously documented in human are missing.

The ligands and biological functions of *Adhesion* GPCRs are largely unknown. Orthologs belonging to the CELSR, HE6 and GP123/124/125 subfamilies have only a single representative in *Ciona* making it an interesting model for studying the roles of these GPCRs. The CELSR family of GPCRs in mammals play a possible functional role in the developmental processes of the brain and the peripheral nervous system [59]. Although a search of the *Ciona* EST database did not reveal any analogous distribution of CELSR homologs in neural tissues, this does not necessarily exclude its expression in these regions. The *Ciona* CELSR model showed maximum (~34%) identity with human CELSR2 while displaying a remarkable level of conservation in the N-terminal domain architecture. The *Ciona* CELSR homolog contains 9 cadherin domain repeats, 2 laminin G-like motifs, 1 laminin EGF-like motif, 7 EGF-like calcium binding motifs and the characteristic GPS domain identical to that in human CELSR2.

Among mammalian LEC receptors, LEC1 and LEC3 expression is abundant in the brain tissue [60]. LEC1 binds to  $\alpha$ -latrotoxin and has a role in exocytosis of the synaptic vesicles [61]. The other vertebrate LEC homologs are unable to function as  $\alpha$ -latrotoxin receptors and their physiological roles remain unknown. *Ciona* also has at least three phylogenetically identifiable LEC homologs for which reliable ESTs can be recovered from the gastrula and neurula, cleavage stage embryo and tail-bud embryo developmental stages suggestive of a role for these receptors in developmental processes [Additional data file 2: Sheet 2]. It should be interesting to explore the presence of LEC homologs in neural tissues of *Ciona* to compare biochemical roles of LEC GPCR biology between vertebrates and protochordates. As many as three gene duplication events can be suspected in the *Ciona* GPCR adhesome (Figure 5). *Ciona* can be a very good model where the developmental role of *Adhesion* orthologs can be investigated at the single cell level because the larvae contain approximately 2600 cells and the lineage of the embryonic cells giving rise to different tissue types has been well determined [62].

#### **The cAMP GPCR-like and Methuselah-like GPCR in *Ciona***

The cAMP/Crl GPCRs belong to an ancient family of GPCRs with homologs observed in the protists, fungi and plants as reported in the Pfam database [63]. However, authentic cAMP GPCRs have been characterized only in the protist *Dictyostelium discoideum* [64]. The *Ciona* cAMP GPCR-like sequence shows best identity with predicted hypothetical proteins in *Danio rerio* and *Nematostella vectensis*. The top hit in a search with the *Ciona* cAMP GPCR-like sequence (ci0100132129) in the PRINTS database was the *Dictyostelium* cAMP receptors (CARs) with matches for 4 out of 7 sequence motifs and an E-value of 1.6e-05 [65]. The next best hit was the GABA<sub>B</sub> GPCR of the *Glutamate* family with matches for 2 out of 14 sequence motifs and an E-value of 23. A recent study on the GPCR repertoire in the *Branchiostoma floridae* genome reports two cAMP GPCR-like sequences with limited sequence identities [12]. The ESTs for the Methuselah-like sequence (ci0100150391) were found only in the adult digestive gland suggesting a possible role for this GPCR in digestive processes. [Additional data file 2: sheet 2].

Although the putative non-GRAFS *Ciona* GPCRs ci0100132129 and ci0100150391 cluster with cAMP and Methuselah GPCR families with significant bootstrap support, they share only limited sequence identities (~19–25%) with homologous members of their respective families. Evidently, sequence homology at these levels does not imply functional homology. It is possible that these sequences are highly divergent homologs of the cAMP and Methuselah GPCR families. The recovery of *Ciona* ESTs for

both receptors confirmed that these genes are endogenously expressed. [Additional data file 2: Sheet 2].

#### **GPCR Orthologs of human disease genes**

*Ciona* has representatives of GPCRs that are associated with diseases in humans. Among the 15 monogenic GPCR disease genes known to date, *Ciona* has orthologs of 6 receptors [Table 3]. The motivation for identification of these GPCR orthologs is to facilitate the conception of studies that will have implications for human health. For example, the mammalian CASR gene is known to be expressed in parathyroid glands and its loss in parathyroid gland in mice is implicated in hypercalcemia, hyperparathyroidism and growth retardation [66]. Similarly, monogenic mutations of the PTHR1 gene is known to be involved in chondrodysplasia and multi-organ disorders, while monogenic mutants of Vasopressin receptor (AVPR2) are implicated in nephrogenic diabetes insipidus [67]. EST matches for CASR, PTHR1 and AVPR2 orthologs can be recovered in *Ciona* [Additional data file 2: Sheet 2]. The tunicate can thus prove an excellent model to probe the roles of these genes. Although physiological roles of these receptors may differ between humans and *Ciona*, studying their roles at the biochemical and genetic level in *Ciona* will potentially pave the way for a better understanding of its vertebrate counterparts.

#### **Sequence conservation of *Ciona* PTHRs**

A particular instance of orthology that will provide clues for the physiological roles of ancestral GPCRs is the PTHR orthologs in *Ciona* that share a high sequence identity with both vertebrate PTHR-1 and PTHR-2. PTHR-1-ligand interactions and their mechanism of action in vertebrates are well understood. PTHR-1 is abundantly expressed in bones and kidney and is known to play an important role in mediating PTH regulated mineral ion homeostasis and endochondral bone formation [68]. Mutations in vertebrate PTHRs and ligands are linked to a number of genetic diseases affecting skeletal development and calcium homeostasis [69]. Hence, the discovery of homologs for these receptors in a protochordate system is of special interest. Unlike PTHR-1, which is ubiquitously expressed, PTHR-2 is expressed in brain, testis, placenta and cardiac endothelium. The physiological roles of PTHR-2 however, are not clearly established [68].

A multiple sequence alignment of *Ciona* PTHRs with those of vertebrates reveal several conserved motifs [Additional data file 4]. The sequence conservation across many functionally important residues responsible for ligand binding and downstream G-protein coupling makes it likely that the putative *Ciona* PTHRs can cross-react specifically with mammalian PTH/PTH related peptides and can signal downstream by coupling to G-proteins. Interestingly, *Ciona* PTHR homologs exhibit high sequence conserva-

**Table 3: *Ciona* genes related to Human GPCR associated diseases.**

Human Disease Category	Human GPCRs	Candidate <i>Ciona</i> GPCRs
Autosomal Dominant Hypocalcemia (ADH), Sporadic Hypoparathyroidism, Familial Hypoparathyroidism	CaSR	ci0100130340
Hypogonadotropic hypogonadism (HH)	GNRHR	ci0100133065 ci0100134571 ci0100152622 ci0100153146
Nephrogenic diabetes insipidus (NDI)	AVPR2	ci0100139176
Precocious puberty, male Pseudohermaphroditism	LHCGR	ci0100133821
Chondrodysplasia, multi-organ disorders, metabolic disorders	PTHR1	ci0100139945 ci0100151327
Sporadic basal cell carcinoma, Somatic mutation	SMOH	ci0100150930
Hypertension, Asthma	ADRB2	ci0100130320 ci0100137803

The list of top hits among *Ciona* genes in a BLASTP search against human disease related GPCRs. The list is based on an E-value cut-off of  $\leq 1e^{-20}$  and further verification by phylogenetic analysis. The *Ciona* matches were checked for their alignment with the human proteins over the length of the TM region. Swiss-Prot identifiers are reported for the human GPCRs.

tion across sites known to be associated with Blomstrand's [70-72] and Jansen's syndrome in humans [73-75]. The absence of defined N- and C-terminal regions in *Ciona* PTHR homologs again seem to be a result of incomplete modelling and a clear picture should emerge with further refinement of the genome or through experimental studies. Information from the existing protein models (ci0100139945 and ci0100151327) should prove useful for performing experiments addressing the physiological roles of these receptors in a protochordate and its evolution in subsequent lineages.

## Conclusion

A genome-wide analysis of the repertoire of putative GPCRs in *Ciona intestinalis* was carried out, revealing many intriguing aspects of ascidian GPCR evolution. While as many as 39% of *Ciona* GPCR sequences are orthologous to vertebrate GPCR subfamilies, a substantial number seem to represent invertebrate-chordate related diversification. The observation of orthologs of several GPCRs related to vertebrate heart and neural tissue physiology raises the possibility that genetic studies of *Ciona* GPCRs can be exploited to assess the functions of complementary higher vertebrate genes. The availability of *Ciona* EST hits with matches for 133 of the 169 putative GPCRs will facilitate further studies to elucidate their functional roles. The compact GPCR repertoire in the *Ciona* model also presents fewer complications in terms of functional redundancy and offers an advantage to study the evolution and function of these receptors. In summary, our analysis leads the way towards understanding the invertebrate-chordate GPCR signalling system and helps identify candidate vertebrate GPCR homologs in *Ciona* to be selected for functional studies.

## Methods

### Description of *Ciona intestinalis* sequence data set

The official version of the *Ciona intestinalis* JGI v2.0 database was used as the source for obtaining the complete proteome [76]. The current size of this tunicate assembly is 173 Mb with approximately 94 Mb of the assembly being mapped to chromosomes, while the remaining 45 % of the genome is included in scaffolds. The proteome dataset includes 15,852 proteins and sequences in the current version include both automated and to a lesser extent, manually curated annotations. A CD-HIT [77] was performed on the proteome set with 99% identity cut-off to obtain a non-redundant dataset that is devoid of splice variants, duplicates and polymorphisms. Non-redundancy of the data set was also verified by a manual check.

### Prediction of Transmembrane regions

Transmembrane regions in the *Ciona* proteome were predicted using secondary structure predictions tools like HMMTOP [78], TMHMM [79] and SOSUI [80] with default settings. A majority of known GPCRs acquire maximum coverage in the 7 TM domain prediction analyses [6]. Since the default settings have been known to under-predict or over-predict TM segments, peptides covering the entire range of 6, 7 and 8 TM domains were retrieved using each of the above programs. The predicted sequences recovered from these programs were later clustered and any redundancy in the (6-8) TM data set was removed. This resulted in 645 unique putative GPCR sequences that were analyzed for GPCR specific patterns using a variety of similarity and pattern search tools in a concerted and comprehensive manner. The strategy to identify protein models with 6/7/8 transmembrane regions while missing out on fragmentary or incomplete GPCR models from the JGI *Ciona* proteome allows for a

better handling of the putative GPCR data set for multiple sequence alignment and phylogenetic studies.

#### **Identification of GPCRs using HMMPFAM**

Customized HMM models for all known GPCR families/subfamilies recognized by the GPCRDB [81], were constructed from ClustalW version 1.83 [82] generated multiple alignments using HMMBUILD and HMMCALIBRATE programs of HMMER package version 2.3.2 [83]. HMMPFAM program was used to query a database of custom-built GPCR HMMs and all (6–8) TM *Ciona* queries that returned hits with an E-value better than 0.01 were extracted. The HMM files and seed alignments are available upon request. Default settings were used in all HMMBUILD and HMMCALIBRATE model builds.

#### **Identification of GPCRs using RPS-BLAST**

Position Specific Scoring Matrices (PSSMs) for all GPCR families/subfamilies recognized by the GPCRDB were generated using BLASTPGP program [84]. A collection of these GPCR PSSMs was then formatted using FORMATRPSDB to construct RPS-BLAST databases. All GPCR family specific PSSMs were generated with known GPCR sequences being used in a two-round iteration search against the corresponding GPCR family-specific database obtained from GPCRDB. For example, a CLASS A (*Rhodopsin*) Amine family PSSM is generated by BLASTPGP using a set of known CLASS A Amine GPCRs as a query against a CLASS A amine database with two rounds of iteration (-j option). FORMATRPSDB operations were carried out with default settings. GPCR family or subfamily specific PSSMs are available upon request. The *Ciona* (6–8) TM data set was searched against a database of custom-made GPCR PSSMs by an RPS-BLAST search with a cut-off at  $E = 0.001$  and the best hits were recovered.

#### **Identification of GPCRs using BLASTP**

GPCR sequences were downloaded from the ftp site of GPCRDB and BLASTP [85] similarity comparisons were performed using the *Ciona* (6–8) TM dataset as query with an E-value =  $10^{-12}$ . *Ciona* queries that returned hits better than the cutoff were extracted into a temporary file.

#### **Identification of remote homologs**

In order to identify homologs for receptors that did not identify themselves with any human sequence, we performed PIPEALIGN searches. PIPEALIGN [86] is a web service that enables collection of potentially homologous sequences based on the query through a series of automated and integrative profile based search steps and finally presents the retrieved hits clustered into subfamilies in the context of a multiple sequence alignment (MSA). The MSA cluster results from PIPEALIGN proved to be a good starting point from which closest neighbours could be identified with a quick-fire NJ or ML analysis.

The nearest neighbour (the species adjacent to the query through the least number of internal nodes or in the case of a tie, *via* the shortest branch length) from such an analysis was identified and hand picked for inclusion into the final analysis.

For example, a typical identification of *Ciona* cAMP GPCR-like and Methuselah-like receptors was carried out by performing searches of these sequence queries on the web-versions of PIPEALIGN, Interproscan [87] and BLASTP (Blastp of Uniref100 database [88]). The hits from these searches were used for a preliminary NJ and ML analysis. After a manual check of the alignment for the hits, the nearest neighbour from such an analysis was then included into the final representative phylogenetic tree reconstruction.

#### **Orthology assignment using BLASTP and phylogenetic methods**

Human GPCR sequences were obtained from Fredriksson et.al [5] as well as from the GPCRDB. *Ciona* protein models that were returned as hits based on similarity/pattern searches (HMM/BLASTP/RPS-BLAST) were compiled together into one dataset. This initial dataset includes many redundant sequences identified simultaneously by two or more similarity/pattern searches. The final non-redundant *Ciona* GPCR dataset consisted of 169 proteins.

To ensure that orthology assignments are sound, a two-way BLASTP search was initially performed followed by phylogenetic analysis. Both *Ciona*-human and human-*Ciona* BLASTP searches were carried out resulting in the identification of sequences that were the top hits to each other. Orthologs were also detected based on cross-genome phylogenetic analysis of the human-*Ciona* GPCR datasets. The final orthology assignment was arrived at based on a consensus of the BLASTP/Reverse BLASTP results and the cross-genome phylogenetic analysis [Additional data file 2: sheet 1]. This final orthology assignment was also independently verified by searches of the *Ciona* sequences against conserved domain databases (Interproscan) and stand-alone local GPCR HMM databases.

#### **Identification of genomic positions of GPCRs using TBLASTN**

The best *Ciona* genome positions were identified for each *Ciona* GPCR using TBLASTN with an E-value cut off at  $10^{-5}$ . All hits were manually inspected and the best genomic positions providing entire coverage of each query was extracted and imported in Excel sheets. [Additional data file 2: sheet 1]

### EST-hits and tissue/developmental stage based expression data

The entire UniGene database of *Ciona* was downloaded from the ftp site [89] and locally installed. *Ciona* GPCR sequences were queried against the UniGene database using TBLASTN with the E-value set at  $1e^{-15}$ . The identifiers of the EST-hit collection were imported in an excel sheet and categorized based on their derivative developmental stage/tissue source to assist in gene finding and transcriptome analysis. [Additional data file 2: sheet 2]

### Phylogenetic Methods

Based on the initial *Ciona*-human BLASTP searches, *Ciona* GPCRs were broadly divided into *Rhodopsin* families and subfamilies and non-*Rhodopsin* receptors. The broadly separated GPCR families were then combined with their related GPCRs from either human or other species. MSAs were then performed using MAFFT version 5 [90] with protein weight matrix JTT200 and employing the E-INS-I strategy with a gap opening penalty of 1.53 and offset value of 0.123. All phylogenetic analyses shown arise out of consideration of only the terminally truncated TM spanning regions. Deviant branches were analysed and the corresponding sequences checked for false positives by looking for GPCR specific signatures or through Interproscan searches. The GRAFS (*Glutamate, Rhodopsin, Adhesion, Frizzled, Secretin*) dataset, the GPCR adhesome, the cAMP GPCR analysis and the non-(LDLRR-GPCR/LGR-like) *Rhodopsin* subsets were bootstrapped 1000 times using SEQBOOT from the PHYLIP package [version] [91]. The alignments were then used for subsequent distance matrices (JTT matrix) calculation by PROTDIST to obtain 1000 matrices. To obtain unrooted trees, neighbour-joining method was employed (NEIGHBOR) and a consensus of 1000 neighbour trees was arrived at using CONSENSE. The topology and phylogenetic support for the above mentioned tree representations was also verified using TREE-PUZZLE analysis. The divergent "*Unclassified/Other*" GPCRs were removed from the final data set to avoid tree artefacts [92].

Maximum likelihood tree of the LGR-like/LDLRR-GPCR cluster was obtained after performing 10,000 quartet puzzling steps with TREE-PUZZLE [version 5.2] [93] using a percentage of invariant sites to be estimated and eight rate categories to describe among-site rate variation with the shape parameter estimated from the data and employing VT (Mueller-Vingron 2000) substitution model. The NJ trees were visualized and edited using SPLITTREE version 4 [94] and the TREE-PUZZLE trees were visualized using Tree view [95]

### Abbreviations

AVPR, Arginine-Vasopressin receptor; BAI, Brain Angiogenesis Inhibitor; CASR, Calcium Sensing Receptor; CAL-

CRL, Calcitonin gene related peptide Receptor; CALCR, Calcitonin Receptor; CELSR, Cadherin EGF LAG Seven G-type Receptors; CNR, Cannabinoid Receptor; CRHR, Corticotropin Releasing Hormone Receptors; CrI, cAMP GPCR-like; CHEM, Chemokine receptor; EDG, Endothelial Differentiation GPCR; EST, Expressed sequence tag; EGF, Epidermal Growth Factor; ETL, EGF-Latrophilin 7 TM domain; FZD, *Frizzled*; GLHR, Glycoprotein hormone Receptor; GPCR, G protein-coupled receptor; GLR, *Glutamate* receptor; GLPR, Glucagon-like peptide Receptor; GNRHR, Gonadotropin Releasing Hormone Receptor; GCGR, Glucagon Receptor; GIPR, Gastric Inhibitory Polypeptide Receptor; GPS, GPCR Proteolytic site domain; HMM, Hidden Markov Model; LDLA, Low Density Lipoprotein Receptor ClassA domain; LDLRR, Low Density Lipoprotein Receptor repeats; LEC, Lectomedin Receptors; LGR, Leucine Rich Repeat containing GPCR; LRP, LDL Receptor related Proteins; LRR, Leucine Rich Repeat domain; LHCGR, Lutenizing hormone/choriogonadotropin receptor; LUTR, Lung 7TM Receptor; MAS, MAS1 oncogene receptor; MCR, Melanocortin Receptors; MRG, MAS-related receptor; MCHR, Melanin Concentrating Hormone Receptor; MECA, Melanocortin, Endothelial Differentiation, Cannabinoid, Adenosine GPCR cluster; MSA, Multiple Sequence Alignment; NJ, Neighbour Joining tree reconstruction method; PUR, Purine; PACAP, Pituitary Adenylyl Cyclase Activating Protein; PSSM, Position Specific Scoring Matrix; PTGER, Prostaglandin Receptor; PTHR, Parathyroid Hormone Receptor; RAIG, Retinoic acid inducible GPCR; RTK, Receptor Tyrosine Kinase; SALPR, Somatostatin and Angiotensin-like peptide receptor; SCTR, *Secretin*; SOG, Somatostatin, Opioid, Galanin cluster; SMOH, Smoothened; TAS, Taste Receptors; TM, Terminally truncated transmembrane spanning segments; VLGR, Very Large G-Protein-Coupled Receptor; VIPR2, Vasoactive intestinal peptide Receptor;

### Authors' contributions

NK carried out the work and has written the first draft of the paper. GKA participated in discussions of the study. NM conceived the study, participated in the analysis and coordination and helped drafting of the final manuscript.

### Additional material

#### Additional file 1

Protein sequences of all *Ciona* GPCRs identified in this study in FASTA format

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-129-S1.pdf>]



**Additional file 2**

Ciona GPCRs with homologs in Humans and other genomes (Sheet 1) and EST-HITS of Ciona GPCRs sorted by developmental stage/tissue (Sheet 2)

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-129-S2.xls>]

**Additional file 3**

Domain architecture of LDLRR-GPCR/LGR-like/GLHR receptor cluster in Ciona

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-129-S3.pdf>]

**Additional file 4**

Multiple sequence alignment of human and rat PTHR with homologs from Ciona. Alignment was generated in MAFFT using gap penalty of 1.53 and offset value 0.123. Scoring matrix is based on JTT200. Colored amino acids are used to refer functionally important residues known in human and rat PTHR1 and their putative homologs in Ciona and human PTHR2. Furthermore transmembrane regions are colored and sequence conservation and similarity represented symbolically.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-129-S4.pdf>]

**Additional file 5**

The GRAFSCM (Glutamate, Rhodopsin, Adhesion, Frizzled, Secretin, cAMP GPCR-like, Methuselah-like) phylogenetic tree in standard Newick format.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-129-S5.tre>]

**Additional file 6**

The non-(LDLRR-GPCR/LGR) Rhodopsin phylogenetic tree in standard Newick format

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-129-S6.tre>]

**Acknowledgements**

We thank the Joint Genome Institute (JGI) for public availability of sequencing data. These sequence data were produced by the US Department of Energy JGI (USA). We also thank IIT Madras and the Bioinformatics Infrastructure Facility, supported by the DBT, Govt of India, for infrastructural support.

**References**

- Chen JY, Huang DY, Peng QQ, Chi HM, Wang XQ, Feng M: **The first tunicate from the Early Cambrian of South China.** *Proc Natl Acad Sci USA* 2003, **100**:8314-8318.
- Delsuc F, Brinkmann H, Chourrout D, Philippe H: **Tunicates and not cephalochordates are the closest living relatives of vertebrates.** *Nature* 2006, **439**:965-968.
- Passamaneck YJ, Di Gregorio A: **Ciona intestinalis: chordate development made simple.** *Dev Dyn* 2005, **233**:1-19.
- Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, et al.: **The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins.** *Science* 2002, **298**:2157-2167.
- Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB: **The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints.** *Mol Pharmacol* 2003, **63**:1256-1272.
- Metpally RP, Sowdhamini R: **Genome wide survey of G protein-coupled receptors in Tetraodon nigroviridis.** *BMC Evol Biol* 2005, **5**:41.
- Hill CA, Fox AN, Pitts RJ, Kent LB, Tan PL, Chrystal MA, Cravchik A, Collins FH, Robertson HM, Zwiebel LJ: **G protein-coupled receptors in Anopheles gambiae.** *Science* 2002, **298**:176-178.
- Lagerstrom MC, Hellstrom AR, Gloriam DE, Larsson TP, Schioth HB, Fredriksson R: **The G protein-coupled receptor subset of the chicken genome.** *PLoS Comput Biol* 2006, **2**:e54.
- Gloriam DE, Fredriksson R, Schioth HB: **The G protein-coupled receptor subset of the rat genome.** *BMC Genomics* 2007, **8**:338.
- Bjarnadottir TK, Gloriam DE, Hellstrand SH, Kristiansson H, Fredriksson R, Schioth HB: **Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse.** *Genomics* 2006, **88**:263-273.
- Metpally RP, Sowdhamini R: **Cross genome phylogenetic analysis of human and Drosophila G protein-coupled receptors: application to functional annotation of orphan receptors.** *BMC Genomics* 2005, **6**:106.
- Nordstrom K, Fredriksson R, Schioth H: **The Branchiostoma genome contains a highly diversified set of G protein-coupled receptors.** *BMC Evol Biol* 2008, **8**(1):9.
- Fredriksson R, Schioth HB: **The repertoire of G-protein-coupled receptors in fully sequenced genomes.** *Mol Pharmacol* 2005, **67**:1414-1425.
- Schioth HB, Nordstrom KJ, Fredriksson R: **Mining the gene repertoire and ESTs for G protein-coupled receptors with evolutionary perspective.** *Acta Physiol (Oxf)* 2007, **190**:21-31.
- Herz JM, Thomsen WJ, Yarbrough GG: **Molecular approaches to receptors as targets for drug discovery.** *J Recept Signal Transduct Res* 1997, **17**:671-776.
- Lichtarge O, Bourne HR, Cohen FE: **An evolutionary trace method defines binding surfaces common to protein families.** *J Mol Biol* 1996, **257**:342-358.
- Flower DR: **Modelling G-protein-coupled receptors for drug design.** *Biochim Biophys Acta* 1999, **1422**:207-234.
- Drews J: **Drug discovery: a historical perspective.** *Science* 2000, **287**:1960-1964.
- Jacoby E, Bouhelal R, Gerspacher M, Seuwen K: **The 7 TM G-protein-coupled receptor target family.** *ChemMedChem* 2006, **1**:761-782.
- Romer AS: *The Vertebrate Body* 4th edition. Philadelphia: W.B.Saunders; 1970.
- Weichert CK: *Elements of Chordate Anatomy* New York: McGraw-Hill; 1953.
- Bathgate RA, Samuel CS, Burazin TC, Gundlach AL, Tregear GW: **Relaxin: new peptides, receptors and novel actions.** *Trends Endocrinol Metab* 2003, **14**:207-213.
- Tensen CP, Van Kesteren ER, Planta RJ, Cox KJ, Burke JF, van Heerikhuizen H, Vreugdenhil E: **A G protein-coupled receptor with low density lipoprotein-binding motifs suggests a role for lipoproteins in G-linked signal transduction.** *Proc Natl Acad Sci USA* 1994, **91**:4816-4820.
- Higgins GA, Miczek KA: **Glutamate receptor subtypes: promising new pharmacotherapeutic targets.** *Psychopharmacology (Berl)* 2005, **179**:1-3.
- Cardoso JC, Pinto VC, Vieira FA, Clark MS, Power DM: **Evolution of Secretin family GPCR members in the metazoa.** *BMC Evol Biol* 2006, **6**:108.
- Huang HC, Klein PS: **The Frizzled family: receptors for multiple signal transduction pathways.** *Genome Biol* 2004, **5**:234.
- Vinson CR, Conover S, Adler PN: **A Drosophila tissue polarity locus encodes a protein containing seven potential transmembrane domains.** *Nature* 1989, **338**:263-264.
- Umbhauer M, Djiane A, Goisset C, Penzo-Mendez A, Riou JF, Boucaut JC, Shi DL: **The C-terminal cytoplasmic Lys-thr-X-X-X-Trp motif in Frizzled receptors mediates Wnt/beta-catenin signalling.** *EMBO J* 2000, **19**:4944-4954.
- Bjarnadottir TK, Fredriksson R, Hoglund PJ, Gloriam DE, Lagerstrom MC, Schioth HB: **The human and mouse repertoire of the**

- Adhesion family of G-protein-coupled receptors.** *Genomics* 2004, **84**:23-33.
30. Klein PS, Sun TJ, Saxe CL 3rd, Kimmel AR, Johnson RL, Devreotes PN: **A chemoattractant receptor controls development in Dictyostelium discoideum.** *Science* 1988, **241**:1467-1472.
  31. Saxe CL 3rd, Ginsburg GT, Louis JM, Johnson R, Devreotes PN, Kimmel AR: **CAR2, a prestalk cAMP receptor required for normal tip formation and late development of Dictyostelium discoideum.** *Genes Dev* 1993, **7**:262-272.
  32. Johnson RL, Saxe CL 3rd, Gollop R, Kimmel AR, Devreotes PN: **Identification and targeted gene disruption of cAR3, a cAMP receptor subtype expressed during multicellular stages of Dictyostelium development.** *Genes Dev* 1993, **7**:273-282.
  33. Raisley B, Zhang M, Hereld D, Hadwiger JA: **A cAMP receptor-like G protein-coupled receptor with roles in growth regulation and development.** *Dev Biol* 2004, **265**:433-445.
  34. Lin YJ, Seroude L, Benzer S: **Extended life-span and stress resistance in the Drosophila mutant methuselah.** *Science* 1998, **282**:943-946.
  35. DeVries ME, Kelvin AA, Xu L, Ran L, Robinson J, Kelvin DJ: **Defining the origins and evolution of the chemokine/chemokine receptor system.** *J Immunol* 2006, **176**:401-415.
  36. Narumiya S, Sugimoto Y, Ushikubi F: **Prostanoid receptors: structures, properties, and functions.** *Physiol Rev* 1999, **79**:1193-1226.
  37. Satake H, Ogasawara M, Kawada T, Masuda K, Aoyama M, Minakata H, Chiba T, Metoki H, Satou Y, Satoh N: **Tachykinin and tachykinin receptor of an ascidian, Ciona intestinalis: evolutionary origin of the vertebrate tachykinin family.** *J Biol Chem* 2004, **279**:53798-53805.
  38. Tello JA, Rivier JE, Sherwood NM: **Tunicate gonadotropin-releasing hormone (GnRH) peptides selectively activate Ciona intestinalis GnRH receptors and the green monkey type II GnRH receptor.** *Endocrinology* 2005, **146**:4061-4073.
  39. Staubli F, Jorgensen TJ, Cazzamali G, Williamson M, Lenz C, Sondergaard L, Roepstorff P, Grimmelikhuijzen CJ: **Molecular identification of the insect adipokinetic hormone receptors.** *Proc Natl Acad Sci USA* 2002, **99**:3446-3451.
  40. Beer MS, Stanton JA, Salim K, Rigby M, Heavens RP, Smith D, McAllister G: **EDG receptors as a therapeutic target in the nervous system.** *Ann N Y Acad Sci* 2000, **905**:118-31.
  41. Toman RE, Milstien S, Spiegel S: **Sphingosine-1-phosphate: an emerging therapeutic target.** *Expert Opin Ther Targets* 2001, **5**:109-123.
  42. Gloriam DE, Bjarnadottir TK, Yan YL, Postlethwait JH, Schioth HB, Fredriksson R: **The repertoire of trace amine G-protein-coupled receptors: large expansion in zebrafish.** *Mol Phylogenet Evol* 2005, **35**:470-482.
  43. Gilbert W: **Why genes in pieces?** *Nature* 1978, **271**:501.
  44. Gilbert W: **Genes-in-pieces revisited.** *Science* 1985, **228**:823-824.
  45. Halls ML, Bond CP, Sudo S, Kumagai J, Ferraro T, Layfield S, Bathgate RA, Summers RJ: **Multiple binding sites revealed by interaction of relaxin family peptides with native and chimeric relaxin family peptide receptors 1 and 2 (LGR7 and LGR8).** *J Pharmacol Exp Ther* 2005, **313**:677-687.
  46. Scott DJ, Layfield S, Yan Y, Sudo S, Hsueh AJ, Tregear GW, Bathgate RA: **Characterization of novel splice variants of LGR7 and LGR8 reveals that receptor signaling is mediated by their unique low density lipoprotein class A modules.** *J Biol Chem* 2006, **281**:34942-34954.
  47. Kobe B, Deisenhofer J: **Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats.** *Nature* 1993, **366**:751-756.
  48. Esser V, Limbird LE, Brown MS, Goldstein JL, Russell DW: **Mutational analysis of the ligand binding domain of the low density lipoprotein receptor.** *J Biol Chem* 1988, **263**:13282-13290.
  49. Russell DW, Brown MS, Goldstein JL: **Different combinations of cysteine-rich repeats mediate binding of low density lipoprotein receptor to two different proteins.** *J Biol Chem* 1989, **264**:21682-21688.
  50. Sudhof TC, Goldstein JL, Brown MS, Russell DW: **The LDL receptor gene: a mosaic of exons shared with different proteins.** *Science* 1985, **228**:815-822.
  51. Yamamoto T, Davis CG, Brown MS, Schneider WJ, Casey ML, Goldstein JL, Russell DW: **The human LDL receptor: a cysteine-rich protein with multiple Alu sequences in its mRNA.** *Cell* 1984, **39**:27-38.
  52. Herz J, Hamann U, Rogne S, Myklebost O, Gausepohl H, Stanley KK: **Surface location and high affinity for calcium of a 500-kd liver membrane protein closely related to the LDL-receptor suggest a physiological role as lipoprotein receptor.** *EMBO J* 1988, **7**:4119-4127.
  53. Nykjaer A, Petersen CM, Moller B, Jensen PH, Moestrup SK, Holtet TL, Etzerodt M, Thogersen HC, Munch M, Andreassen PA, Gliemann J: **Purified alpha 2-macroglobulin receptor/LDL receptor-related protein binds urokinase plasminogen activator inhibitor type-I complex. Evidence that the alpha 2-macroglobulin receptor mediates cellular degradation of urokinase receptor-bound complexes.** *J Biol Chem* 1992, **267**:14543-14546.
  54. Herz J, Couthier DE, Hammer RE: **Correction: LDL receptor-related protein internalizes and degrades uPA-PAI-1 complexes and is essential for embryo implantation.** *Cell* 1993, **73**:428.
  55. Sachinidis A, Locher R, Mengden T, Vetter W: **Low-density lipoprotein elevates intracellular calcium and pH in vascular smooth muscle cells and fibroblasts without mediation of LDL receptor.** *Biochem Biophys Res Commun* 1990, **167**:353-359.
  56. Block LH, Knorr M, Vogt E, Locher R, Vetter W, Groscurth P, Qiao BY, Pometta D, James R, Regenass M, Pletscher A: **Low density lipoprotein causes general cellular activation with increased phosphatidylinositol turnover and lipoprotein catabolism.** *Proc Natl Acad Sci USA* 1988, **85**:885-889.
  57. Wu YQ, Jorgensen EV, Handwerker S: **High density lipoproteins stimulate placental lactogen release and adenosine 3',5'-monophosphate (cAMP) production in human trophoblast cells: evidence for cAMP as a second messenger in human placental lactogen release.** *Endocrinology* 1988, **123**:1879-1884.
  58. Olinski RP, Dahlberg C, Thorndyke M, Hallbook F: **Three insulin-relaxin-like genes in Ciona intestinalis.** *Peptides* 2006, **27**:2535-2546.
  59. Formstone CJ, Little PF: **The flamingo-related mouse Celsr family (Celsr1-3) genes exhibit distinct patterns of expression during embryonic development.** *Mech Dev* 2001, **109**:91-94.
  60. Matsushita H, Leliana V, Ushkaryov Y: **The latrophilin family: multiply spliced G protein-coupled receptors with differential tissue distribution.** *FEBS Lett* 1999, **443**(3):348-352.
  61. Bjarnadottir TK, Fredriksson R, Schioth HB: **The Adhesion GPCRs: a unique family of G protein-coupled receptors with important roles in both central and peripheral tissues.** *Cell Mol Life Sci* 2007, **642**:2104-2119.
  62. Kusakabe T, Yoshida R, Kawakami I, Kusakabe R, Mochizuki Y, Yamada L, Shin-i T, Kohara Y, Satoh N, Tsuda M, Satou Y: **Gene expression profiles in tadpole larvae of Ciona intestinalis.** *Dev Biol* 2002, **242**:188-203.
  63. Pfam [<http://www.sanger.ac.uk/Software/Pfam/>]
  64. Prabhu Y, Eichinger L: **The Dictyostelium repertoire of seven transmembrane domain receptors.** *Eur J Cell Biol* 2006, **85**:937-946.
  65. FingerPRINTScan [<http://www.bioinf.manchester.ac.uk/fingerPRINTScan/>]
  66. Ho C, Conner DA, Pollak MR, Ladd DJ, Kifor O, Warren HB, Brown EM, Seidman JG, Seidman CE: **A mouse model of human familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism.** *Nat Genet* 1995, **11**:389-394.
  67. Insel PA, Tang CM, Hahntow I, Michel MC: **Impact of GPCRs in clinical medicine: monogenic diseases, genetic variants and drug targets.** *Biochim Biophys Acta* 2007, **1768**:994-1005.
  68. Mannstadt M, Juppner H, Gardella TJ: **Receptors for PTH and PTHrP: their biological importance and functional properties.** *Am J Physiol* 1999, **277**:F665-675.
  69. Huch K, Kleffner S, Stove J, Puhl W, Gunther KP, Brenner RE: **PTHrP, PTHr, and FGFR3 are involved in the process of endochondral ossification in human osteophytes.** *Histochem Cell Biol* 2003, **119**:281-287.
  70. Blomstrand S, Claesson I, Save-Soderbergh J: **A case of lethal congenital dwarfism with accelerated skeletal maturation.** *Pediatr Radiol* 1985, **15**:141-143.
  71. Karaplis AC, He B, Nguyen MT, Young ID, Semeraro D, Ozawa H, Amizuka N: **Inactivating mutation in the human parathyroid hormone receptor type I gene in Blomstrand chondrodysplasia.** *Endocrinology* 1998, **139**:5255-5258.
  72. Zhang P, Jobert AS, Couvineau A, Silve C: **A homozygous inactivating mutation in the parathyroid hormone/parathyroid**

- hormone-related peptide receptor causing Blomstrand chondrodysplasia. *J Clin Endocrinol Metab* 1998, **83**:3365-3368.
73. Schipani E, Kruse K, Juppner H: **A constitutively active mutant PTH-PTHrP receptor in Jansen-type metaphyseal chondrodysplasia.** *Science* 1995, **268**:98-100.
  74. Schipani E, Langman C, Hunzelman J, Le Merrer M, Loke KY, Dillon MJ, Silve C, Juppner H: **A novel parathyroid hormone (PTH)/PTH-related peptide receptor mutation in Jansen's metaphyseal chondrodysplasia.** *J Clin Endocrinol Metab* 1999, **84**:3052-3057.
  75. Schipani E, Langman CB, Parfitt AM, Jensen GS, Kikuchi S, Kooh SW, Cole WG, Juppner H: **Constitutively activated receptors for parathyroid hormone and parathyroid hormone-related peptide in Jansen's metaphyseal chondrodysplasia.** *N Engl J Med* 1996, **335**:708-714.
  76. **Ciona intestinalis v2.0** [<http://genome.igi-psf.org/Cioin2/Cioin2.home.html>]
  77. Li W, Jaroszewski L, Godzik A: **Clustering of highly homologous sequences to reduce the size of large protein databases.** *Bioinformatics* 2001, **17**:282-283.
  78. Tusnady GE, Simon I: **The HMMTOP transmembrane topology prediction server.** *Bioinformatics* 2001, **17**:849-850.
  79. Krogh A, Larsson B, von Heijne G, Sonnhammer EL: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** *J Mol Biol* 2001, **305**:567-580.
  80. Hirokawa T, Boon-Chiang S, Mitaku S: **SOSUI: classification and secondary structure prediction system for membrane proteins.** *Bioinformatics* 1998, **14**:378-379.
  81. Horn F, Bettler E, Oliveira L, Campagne F, Cohen FE, Vriend G: **GPCRDB information system for G protein-coupled receptors.** *Nucleic Acids Res* 2003, **31**:294-297.
  82. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD: **Multiple sequence alignment with the Clustal series of programs.** *Nucleic Acids Res* 2003, **31**:3497-3500.
  83. Eddy SR: **Profile hidden Markov models.** *Bioinformatics* 1998, **14**:755-763.
  84. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
  85. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
  86. Plewniak F, Bianchetti L, Brelivet Y, Carles A, Chalmel F, Lecompte O, Mochel T, Moulinier L, Muller A, Muller J, Prigent V, Ripp R, Thierry JC, Thompson JD, Wicker N, Poch O: **PipeAlign: A new toolkit for protein family analysis.** *Nucleic Acids Res* 2003, **31**:3829-3832.
  87. Zdobnov EM, Apweiler R: **InterProScan – an integration platform for the signature-recognition methods in InterPro.** *Bioinformatics* 2001, **17**:847-848.
  88. **UniProt** [<http://www.ebi.uniprot.org/index.shtml>]
  89. **NCBI UniGene** [<ftp://ftp.ncbi.nih.gov/repository/UniGene/>]
  90. Katoh K, Kuma K, Toh H, Miyata T: **MAFFT version 5: improvement in accuracy of multiple sequence alignment.** *Nucleic Acids Res* 2005, **33**:511-518.
  91. Felsenstein J: **PHYLP – Phylogeny Inference Package (Version 3.2).** *Cladistics* 1989, **5**:164-166.
  92. Baldauf SL: **Phylogeny for the faint of heart: a tutorial.** *Trends Genet* 2003, **19**:345-351.
  93. Schmidt HA, Strimmer K, Vingron M, von Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.** *Bioinformatics* 2002, **18**:502-504.
  94. Huson DH, Bryant D: **Application of phylogenetic networks in evolutionary studies.** *Mol Biol Evol* 2006, **23**:254-267.
  95. Page RD: **TreeView: an application to display phylogenetic trees on personal computers.** *Comput Appl Biosci* 1996, **12**:357-358.
  96. Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P: **SMART 5: domains in the context of genomes and networks.** *Nucleic Acids Res* 2006, **34**:D257-260.
  97. **GTPred** [<http://gdds.pharm.kyoto-u.ac.jp/services/gtpred/>]
  98. Gardella TJ, Luck MD, Fan MH, Lee C: **Transmembrane residues of the parathyroid hormone (PTH)/PTH-related peptide receptor that specifically affect binding and signaling by agonist ligands.** *J Biol Chem* 1996, **271**:12820-12825.
  99. Carter PH, Shimizu M, Luck MD, Gardella TJ: **The hydrophobic residues phenylalanine 184 and leucine 187 in the type-I parathyroid hormone (PTH) receptor functionally interact with the amino-terminal portion of PTH-(1-34).** *J Biol Chem* 1999, **274**:31955-31960.
  100. Bergwitz C, Jusseaume SA, Luck MD, Juppner H, Gardella TJ: **Residues in the membrane-spanning and extracellular loop regions of the parathyroid hormone (PTH)-2 receptor determine signaling selectivity for PTH and PTH-related peptide.** *J Biol Chem* 1997, **272**:28861-28868.
  101. Turner PR, Mefford S, Bambino T, Nissenson RA: **Transmembrane residues together with the amino terminus limit the response of the parathyroid hormone (PTH) 2 receptor to PTH-related peptide.** *J Biol Chem* 1998, **273**:3830-3837.
  102. Plati J, Tsomaia N, Piserchio A, Mierke DF: **Structural features of parathyroid hormone receptor coupled to Galpha(s)-protein.** *Biophys J* 2007, **92**:535-540.
  103. Iida-Klein A, Guo J, Takemura M, Drake MT, Potts JT Jr, Abou-Samra A, Bringham FR, Segre GV: **Mutations in the second cytoplasmic loop of the rat parathyroid hormone (PTH)/PTH-related protein receptor result in selective loss of PTH-stimulated phospholipase C activity.** *J Biol Chem* 1997, **272**:6882-6889.
  104. Huang Z, Chen Y, Pratt S, Chen TH, Bambino T, Nissenson RA, Shoback DM: **The N-terminal region of the third intracellular loop of the parathyroid hormone (PTH)/PTH-related peptide receptor is critical for coupling to cAMP and inositol phosphate/Ca2+ signal transduction pathways.** *J Biol Chem* 1996, **271**:33382-33389.
  105. Gardella TJ, Juppner H, Wilson AK, Keutmann HT, Abou-Samra AB, Segre GV, Bringham FR, Potts JT Jr, Nussbaum SR, Kronenberg HM: **Determinants of [Arg2]PTH-(1-34) binding and signaling in the transmembrane region of the parathyroid hormone receptor.** *Endocrinology* 1994, **135**:1186-1194.
  106. Mannstadt M, Luck MD, Gardella TJ, Juppner H: **Evidence for a ligand interaction site at the amino-terminus of the parathyroid hormone (PTH)/PTH-related protein receptor from cross-linking and mutational studies.** *J Biol Chem* 1998, **273**:16890-16896.
  107. Lee C, Luck MD, Juppner H, Potts JT Jr, Kronenberg HM, Gardella TJ: **Homolog-scanning mutagenesis of the parathyroid hormone (PTH) receptor reveals PTH-(1-34) binding determinants in the third extracellular loop.** *Mol Endocrinol* 1995, **9**:1269-1278.
  108. Gensure RC, Shimizu N, Tsang J, Gardella TJ: **Identification of a contact site for residue 19 of parathyroid hormone (PTH) and PTH-related protein analogs in transmembrane domain two of the type I PTH receptor.** *Mol Endocrinol* 2003, **17**:2647-2658.
  109. Adams AE, Bisello A, Chorev M, Rosenblatt M, Suva LJ: **Arginine 186 in the extracellular N-terminal region of the human parathyroid hormone 1 receptor is essential for contact with position 13 of the hormone.** *Mol Endocrinol* 1998, **12**:1673-1683.
  110. Bisello A, Adams AE, Mierke DF, Pellegrini M, Rosenblatt M, Suva LJ, Chorev M: **Parathyroid hormone-receptor interactions identified directly by photocross-linking and molecular modeling studies.** *J Biol Chem* 1998, **273**:22498-22505.
  111. Greenberg Z, Bisello A, Mierke DF, Rosenblatt M, Chorev M: **Mapping the bimolecular interface of the parathyroid hormone (PTH)-PTH1 receptor complex: spatial proximity between Lys(27) (of the hormone principal binding domain) and leu(261) (of the first extracellular loop) of the human PTH1 receptor.** *Biochemistry* 2000, **39**:8142-8152.
  112. **NetNGlyc 1.0 Server** [<http://www.cbs.dtu.dk/services/NetNGlyc/>]