Molecular Phylogenetics and Evolution 192 (2024) 107986

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Comparative genomic analysis of chemosensory-related gene families in gastropods

Johnma José Rondón [a,b], Vadim A. Pisarenco [c,d], José Ramón Pardos-Blas [e], Alejandro Sánchez-Gracia [c,d], Rafael Zardoya [e,*], Julio Rozas [c,d,*]

[a] Fundación Instituto Leloir, Instituto de Investigaciones Bioquímicas de Buenos Aires (IIBBA-CONICET), Buenos Aires, Argentina
[b] Instituto de Ecología, Genética y Evolución de Buenos Aires (IEGEBA-CONICET) Buenos Aires, Argentina
[c] Departament de Genètica, Microbiologia i Estadística, Universitat de Barcelona (UB), Barcelona, Spain
[d] Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona (UB), Barcelona, Spain
[e] Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain

ARTICLE INFO

Keywords:
Gastropoda
Chemoreceptors
Chemosensory gene families
Gene clusters
GPCRs
WGD

ABSTRACT

Chemoreception is critical for the survival and reproduction of animals. Except for a reduced group of insects and chelicerates, the molecular identity of chemosensory proteins is poorly understood in invertebrates. Gastropoda is the extant mollusk class with the greatest species richness, including marine, freshwater, and terrestrial lineages, and likely, highly diverse chemoreception systems. Here, we performed a comprehensive comparative genome analysis taking advantage of the chromosome-level information of two Gastropoda species, one of which belongs to a lineage that underwent a whole genome duplication event. We identified thousands of previously uncharacterized chemosensory-related genes, the majority of them encoding G protein-coupled receptors (GPCR), mostly organized into clusters distributed across all chromosomes. We also detected gene families encoding degenerin epithelial sodium channels (DEG-ENaC), ionotropic receptors (IR), sensory neuron membrane proteins (SNMP), Niemann–Pick type C2 (NPC2) proteins, and lipocalins, although with a lower number of members. Our phylogenetic analysis of the GPCR gene family across protostomes revealed: (i) remarkable gene family expansions in Gastropoda; (ii) clades including members from all protostomes; and (iii) species-specific clades with a substantial number of receptors. For the first time, we provide new and valuable knowledge into the evolution of the chemosensory gene families in invertebrates other than arthropods.

1. Introduction

The chemosensory system is essential for the survival and reproduction of animals. The olfactory and taste senses are responsible for the detection of chemical signals, a critical process that, at the molecular level, is mediated by proteins that usually belong to medium- to large-sized gene families (Robertson, 2019). Peripheral chemosensory system (CS) functioning relies on two main groups of proteins, membrane receptors and extracellular water-soluble ligand-binding proteins (Sánchez-Gracia et al., 2009). Studies on arthropods and vertebrates show that chemoreceptors are encoded by large gene families (Nei et al., 2008; Vizueta et al., 2020a; Vizueta et al., 2018), with most of them having a seven transmembrane domain (7-TM) structure (Hilger et al., 2018; Krishnan et al., 2014). Some members of these receptor families, however, can be involved in physiological processes other than chemosensation (Böhme & Beck-Sickinger, 2009). Vertebrate genomes encode six different chemosensory receptor families: olfactory receptors (vOR), trace amine-associated receptors (TAAR), and vomeronasal receptors type 1 and 2 (V1R and V2R) (Buck & Axel, 1991; Liberles & Buck, 2006), which are associated with the olfactory system, and type 1 and 2 (T1R and T2R) taste receptors (Adler et al., 2000; Li et al., 2002). All of these families belong to the well known rhodopsin-like (Class A) G protein-coupled receptor (GPCR) superfamily (equivalent to group R of the GRAFS classification (Alexander et al., 2021)). Arthropod chemoreceptor families include the odorant (OR) and gustatory (GR) receptors, ionotropic receptors (IR), a group of highly divergent ionotropic glutamate receptors (iGluR) (Croset et al., 2010), and degenerin amiloride-sensitive ion channels (DEG-ENaC; (Pikielny, 2012)). Despite

* Corresponding authors at: Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain (R. Zardoya). Departament de Genètica, Microbiologia i Estadística, Universitat de Barcelona (UB), Barcelona, Spain (J. Rozas).
E-mail addresses: rafaz@mncn.csic.es (R. Zardoya), jrozas@ub.edu (J. Rozas).

https://doi.org/10.1016/j.ympev.2023.107986
Received 18 July 2023; Received in revised form 24 November 2023; Accepted 7 December 2023
Available online 22 December 2023

exhibiting a 7-TM structure, both vertebrate and insect OR are not homologs; indeed, they have an opposite membrane topology and use different signal transduction pathways, constituting an astonishing case of functional convergence (Benton et al., 2020; Clyne et al., 1999; Derby et al., 2016; Vieira & Rozas, 2011). On the other hand, the small soluble proteins secreted in the aqueous space surrounding insect chemoreceptors include members of the odorant-binding (OBP), chemosensory (CSP), and Niemann–Pick type C2 (NPC2) gene families (Angeli et al., 1999; Pelosi & Maida, 1995; Storch & Xu, 2009). The candidate carrier (CCP) gene family can also be added to this last group, although its secretion in the vicinity of chemoreceptors has not been confirmed (Vizueta et al., 2017). Once more, despite the same name (and general function), vertebrate and insect OBP families are not homologs; with vertebrate OBP (vOBP) belonging to the lipocalin family.

The independent evolutionary origin of the chemosensory gene families in vertebrates and arthropods prompts for an in-depth characterization of these gene families in other invertebrate taxa, particularly taking into account that most invertebrates have an aquatic lifestyle. Despite the fact that few studies have investigated the correlation between the type of habitat and the molecular mechanisms of chemoreception, the role of the environment in the evolution of chemosensory gene families is evident. With 40,000–90,000 estimated species, Gastropoda is the extant mollusk class with the greatest species richness, showing an astonishing morphological and ecological diversity (Albano, 2021). This group includes marine, freshwater, and terrestrial species with very different life histories, in which chemoreception surely plays a fundamental role in its ecological diversification. Therefore, gastropods could be a sound model system to characterize the wider diversity of chemosensory-related genes outside insects and chelicerates. Thus far, the few attempts to explore such diversity focused on the detection of differentially expressed genes of the GPCR and iGluR families in the rhinophore and oral tentacles of sea hares of the genus *Aplysia* (infraclass Heterobranchia) (Croset et al., 2010; Cummins et al., 2009). Yet, there are no wide genomic surveys of the complete repertoire of chemosensory-related genes reported in gastropods, as high quality (chromosome-level) genomes for this group have not been available until recently. Within Gastropoda, 60% of the species diversity is concentrated into infraclass Caenogastropoda, which includes mostly marine but also estuarine, freshwater, and terrestrial lineages (Albano, 2021). Early divergent caenogastropod lineages include mostly detritivore snails that browse on micro-organisms and decomposing organic matter. During the evolutionary history of Caenogastropoda, other dietary specializations were acquired including herbivory, suspension-feeding, and carnivory (Albano, 2021). In particular, the order Neogastropoda includes active carnivores, which have evolved highly sophisticated predatory systems (Lemarcis et al., 2022).

In this context, we compared the chromosome-level genomes of two caenogastropod species, the golden apple snail *Pomacea canaliculata* (Lamarck, 1822), belonging to family Ampullariidae (order Ampullariida), and the Mediterranean cone snail *Lautoconus ventricosus* (Gmelin, 1791), belonging to family Conidae (order Neogastropoda), with the aim of describing for the first time the main gene families related to the chemosensory system of gastropods (Fig. 1). The golden apple snail lives in freshwater streams and ponds, feeds mostly on organic matter and macrophytes (Lach et al., 2000), and has both gills and a lung to breathe air (Mueck et al., 2020). Native to South America, *P. canaliculata* is a highly invasive species in southern and eastern Asia (Hayes et al., 2008) and North America (Rawlings et al., 2007), where it causes serious alterations of native wetlands. The Mediterranean cone *L. ventricosus* is a marine venomous snail endemic to the Mediterranean Sea and adjacent Atlantic waters (Abalde et al., 2020). It lives on rocky shores in the intertidal zone, feeding actively on worms, which are captured through the injection of a cocktail of potent neurotoxic peptides termed conotoxins. Despite the absence of a chromosome-level assembly, we also included genome data from the sea hare *Aplysia californica* (Cooper, 1865) in the comparative analyses, as it is the only gastropod species with transcriptomic information from the chemosensory organs reported (Cummins et al., 2009). Our analysis using our bioinformatic pipeline BITACORA (Vizueta et al., 2020b) allowed identifying thousands of GPCR members, most of them organized into large gene clusters. We also identified members of other chemosensory-related gene families, but with much more modest repertory sizes. Our findings,

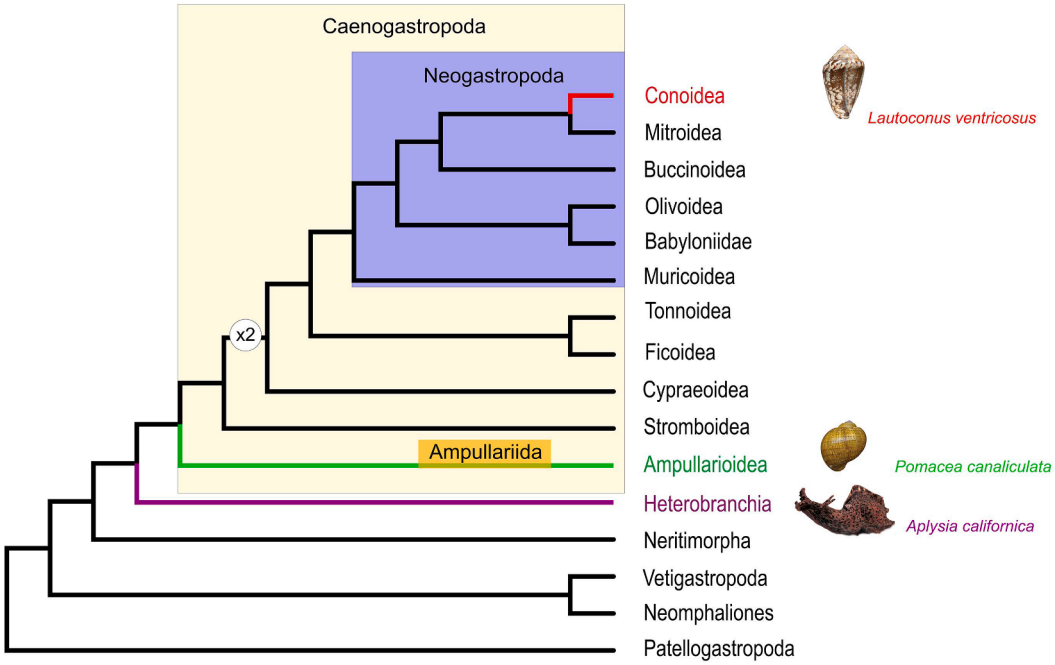

Fig. 1. Gastropod phylogeny showing the relationships among the species analyzed in this work as well as the WGD event (x2) close to the ancestor of the Neogastropoda lineage (Hallinan & Lindberg, 2011; Liu et al., 2021; Pardos-Blas et al., 2021). Phylogenetic relationships among major gastropod lineages were based on (Uribe et al., 2019), caenogastropod lineages were based on (Lemarcis et al., 2022), except for the position of *P. canaliculata*, which was based on (Albano, 2021; Osca et al., 2015). Picture from *P. canaliculata* was taken by H. Zell.

which are examined in the context of the ancient WGD that occurred within the caenogastropod lineage, contribute novel and valuable knowledge into the evolution of chemosensory gene families in invertebrates other than arthropods.

2. Materials and methods

2.1. Identification of chemosensory genes in gastropod genomes

Whole-genome sequences and annotation files of *L. ventricosus* (BioProject PRJNA678883; (Pardos-Blas et al., 2021)), *P. canaliculata* (BioProject PRJNA427478; (Liu et al., 2018)), and *A. californica* (BioProject PRJNA13635; (Knudsen et al., 2006)) were downloaded from NCBI (National Center for Biotechnology Information; (Sayers et al., 2021)) (Table 1). In the case of *L. ventricosus* we also compared its two available assemblies, the first consists of a total of 35 chromosomes (chromosomal assembly) and for the other, 19,364 unplaced scaffolds were added (non-chromosomal assembly). For the other two gastropods, we used the only assembly available for each species, chromosomal for *P. canaliculata* and non-chromosomal for *A. californica.* We applied the BITACORA v.1.2.1 bioinformatic pipeline (Vizueta et al., 2020b) in two successive search rounds using as query a set of previously identified chemosensory proteins from invertebrate lineages phylogenetically close to gastropods (other mollusks, as well as other more distant ones such as arthropods, and nematodes). The chemosensory sequences downloaded from NCBI were used to create a local reference database of proteins for each family, including GPCRs, DEG-ENaC, IR/iGluR, GR, OR, OBP, CSP, CCP, NPC2, SNMP, and lipocalin protein families (see databases folder in the supplementary data). We used HMMER v3.3 (Eddy, 2011) to build the HMM (hidden Markov model) profile for each of these families. The sequences obtained in the first search round were used to improve HMM profiles for the second search round. Candidate sequences were classified as members of one of the query gene families based on a cutoff E-value of 1×10^{-8} in BLAST searches (Camacho et al., 2009).

Protein sequences belonging to GPCR subfamilies were used to search for GPCR candidates, including the 7-TM receptor rhodopsin (7TM_1; ID Pfam database PF00001), *frizzled* (PF01534), 7TM_2 secretin (PF00002), 7TM_GPCR_Srg (PF02118), 7TM_GPCR_Srsx (PF10320), 7TM_GPCR_Srv (PF10323), 7TM_GPCR_Srw (PF10324) and 7TM_GPCR_Srx (PF10328). We also included GPCR sequences (subfamilies A, B, and C) reported in *A. californica* (Cummins et al., 2009). To classify GPCR encoding candidates, we searched for functional domains with the Pfam and Superfamily databases (Gough et al., 2001; Mistry et al., 2021).

Gene models were further validated by inspecting the transmembrane domain (TMD) typical of GPCR and the protein domains characteristic of the DEG-ENaC and IR/iGluR families, using TMHMM 2.0 (Krogh et al., 2001) and InterProScan v.5.56-89.0 (Jones et al., 2014) with the Pfam and the Conserved Domain Database (CDD) databases, respectively. To identify members of the IR/iGluR families, we used the Pfam domains PF10613, PF00060, and PF01094, whereas for DEG-ENaC, we used the Pfam domain PF00858. We classified the novel sequences into three categories based on the functional and structural criteria as described in (Vizueta et al., 2018). Briefly, these categories included (i) sequences encoding complete proteins (*Scp* set): for the GPCR family, sequences with six or more TMDs were included; (ii) The *Smin* set (the minimum number of sequences), those encoding complete (*Scp* set) plus some incomplete proteins that either encoded at least five TMDs (for GPCRs) or showed more than 80% of the corresponding average protein domain length. This value, which represents the minimum number of potentially functional sequences was used for comparative purposes in the GPCR, IR/iGluR, and DEG-ENaC families, across the different genomes, and (iii) all sequences identified by BITACORA (*Smax* set), which represent the maximum possible number of members from a particular gene family. In summary, our classification considers structural aspects in the sequences (gene completeness) to infer their functional character and thus, include them into a given category.

2.2. Identification of chemosensory-related gene clusters in the genomes of *L. ventricosus* and *P. canaliculata*

The chromosomal distribution of chemosensory-related genes was investigated in the two species with chromosome-level genomes (*L. ventricosus* and *P. canaliculata*). We extracted the genome coordinates of chemosensory-related genes in the described chromosomes of the surveyed species (35 and 14 for *L. ventricosus* and *P. canaliculata,* respectively) from GFF files output from BITACORA. We identified the genomic clusters for each chemosensory family following the approach of Vieira et al., 2007 and with the adjustments presented in Escuer et al., 2022:

$$C_L = g(n - 1),$$

where C_L is the maximum length of a genomic region containing n genes of a particular chemosensory family to be considered a cluster, with g being the maximum physical distance between two genes to be considered as clustered. If two or more genes were located at a distance less than C_L, they were considered members of the same gene cluster (Escuer et al., 2022). Since *L. ventricosus* and *P. canaliculata* greatly differ in their genome size, we used different g values to compute C_L. For simplicity, we used the same C_L value for all chemosensory families in the same genome; specifically, C_L values were estimated using information from the largest gene family (GPCR), which corresponds to the more conservative scenario avoiding the overestimation of the number of clusters in families with a small number of members.

2.3. Phylogenetic analyses and evolutionary distances

We aligned candidate protein sequences from each family with MAFFT v.7.453 (Katoh & Standley, 2013), and misaligned regions were trimmed with trimAl v.14 (using the options *-gt* 0.2, *-st* 0.0001, and *-cons* 60) (Capella-Gutiérrez et al., 2009). The alignment strategy for each gene family was chosen by MAFFT run in *-auto* mode. In particular for the GPCR family we used the FFT-NS-i algorithm due to the number of sequences analysed. The IR/iGluR and GPCR sequences were first aligned using only full (*Scp*) sequences; then partial proteins were added with the *-addfragments* and *-keeplength* options in MAFFT. For the DEG-ENaC family, since a moderate number of members were found, we decided to include all the sequences identified in the analysis (*Smax*). Gene family phylogenetic trees were built using a maximum likelihood

Table 1
Genomic statistics.

	L. ventricosus		*P. canaliculata*	*A. californica*
	Whole assembly	Chromosomic assembly	Whole/Chromosomic assembly	Whole assembly
Assembly size (bp)	3,592,060,885	3,154,414,418	440,159,624	927,310,431
Number of scaffolds	19,399	35	24 (14)*	4,332
Protein-coding genes	31,591	28,579	21,533	26,663

* Number of pseudochromosomes in parenthesis.

(ML) approach in IQTREE v.2.1.2 (Minh et al., 2020), which also allowed us to estimate the substitution model with the best fit to each dataset. The support values of each branch were obtained after 1,000 bootstrap replications. In the GPCR tree, the rooting was placed in the branch connecting the clearly differentiated clade with GPCR sequences of *A. californica* with the sequences of the other gastropods. The ML phylogeny of invertebrate GPCRs was constructed using the sequences from gastropods identified in this work, together with homologous sequences belonging to GPCR families from Nematoda (*Caenorhabditis elegans*), Mollusca: Gastropoda (*Lottia gigantea*), Arthropoda (*Anopheles gambiae, Drosophila melanogaster, Daphnia pulex, Pediculus humanus*), Platyhelminthes (*Schistosoma mansoni* and *Schmidtea mediterranea*), and Cnidaria (*Nematostella vectensis*) (Krishnan et al., 2014; Saberi et al., 2016). We also incorporated sequences of *Dictyostelium fasciculatum* as outgroups (Krishnan et al., 2014). We used the iTOL web tool (Letunic & Bork, 2021) to customize tree visualization by adding information about the species and chromosomal location of each sequence. We calculated the pairwise physical and evolutionary distances between GPCR paralogs as in Escuer et al., 2022. Evolutionary distances were estimated using the MEGA-CC v.11.0.11 (Kumar et al., 2012) and the JTT-F substitution model (Jones et al., 1992) with gamma-distributed heterogeneous rate variation among sites and 14 discrete classes (alpha = 7.37), found by IQ-TREE as the best-fit model for the evolution of this family in invertebrates.

3. Results

3.1. GPCRs are the most abundant chemosensory-related gene families in the surveyed gastropod genomes

First, regarding the number of putative genes for each chemosensory family, we found that by far the GPCR gene family is the most abundant gene family across the three gastropods studied. The cone snail *L. ventricosus* could encode 2102 and 2053 members (*Smin* values) upon the non-chromosomal, and chromosome-level assemblies, respectively (Tables 2, S1 and S2). The other gene families encompassed a much lower number of members (51 IR/iGluR, 18 DEG-ENaC, and less than 10 copies of the NPC2, SNMP, and lipocalin families; Tables S1 and S2). The genome of *P. canaliculata* also encodes a large number of chemosensory-related genes, about half that of *L. ventricosus i.e.*, 1222 and 32 genes (*Smin*) of the GPCR and IR/iGluR families, respectively (Tables 1 and S2) in the chromosome assembly. In the *Aplysia* genome, we found fewer GPCR than in *P. canaliculata* (872 copies), and a similar number of other gene families. In the surveyed genomes of the three species, the number of genes identified by our pipeline was higher than that reported in the annotations released for each assembly (Tables S1–S3).

Second, regarding the classification of the putative GPCR genes found in each gastropod genome, the structural and functional annotation revealed that 95.3% (2881 members in *L. ventricosus*), 98.6% (1406 members in *P. canaliculata*), and 95.3% (1064 members in *A. californica*) of the sequences belonged to class A of the GPCR (SSF81321) based on the superfamily database analysis (Table S3). The most represented functional Pfam domains were the 7-TM receptors (PF00001); 49.6% members in *L. ventricosus*; 53.7% in *P. canaliculata*; and 46.8% in *A. californica*. The second most represented domain was the Serpentine type 7-TM GPCR chemoreceptor Srw (PF10324) (Table S3 and Figure S1). As expected, we found that the number of GPCRs encoded in the high-quality genomes was much higher (at least 10-fold higher) than those reported in *A. californica* and *L. gigantea* (Cummins et al., 2009; Robertson, 2015).

3.2. Chemosensory-related genes are mostly organized into genomic clusters in the genomes of gastropods

The chromosome-level genome assemblies allowed performing a comprehensive analysis of the physical location and chromosomal distribution of family members. Overall, we found that, chemosensory-related genes were distributed evenly across all chromosomes in *L. ventricosus* and *P. canaliculata*, with some exceptions, such as chromosome 24 in *L. ventricosus* and chromosome 6 in *P. canaliculata*, which

Table 2
Comparison of the number of genes per family, clusters, and genes within and outside of the genome clusters using different values of *g* in *L. ventricosus* and *P. canaliculata*.

Family	*L. ventricosus* 35 chromosomes assembly								
	Total number of genes, and *Scp* values	*g* = 100 kb				*g* = 150 kb			
		Number of clusters	Genes in clusters	Genes out of clusters	% genes in clusters	Number of clusters	Genes in clusters	Genes out of clusters	% genes in clusters
GPCRs*	3,022 (1,602)*	235	989	613	62%	243	1,127	475	70%
IR	57	13	28	29	49%	13	31	26	54%
DEG-ENaC	18	4	9	9	50%	5	12	6	67%
NPC2	3	0	0	0	-	0	0	0	-
SNMP	7	0	0	0	-	0	0	0	-
Lipocalins	0	0	0	0	-	0	0	0	-
GR	0	0	0	0	-	0	0	0	-
Total	3,107	252	1,026	758	-	256	1,184	600	-
	P. canaliculata 14 chromosomes assembly								
	Total number of genes, and *Scp* values	*g* = 10 kb				*g* = 100 kb			
		Number of clusters	Genes in clusters	Genes out of clusters	% genes in clusters	Number of clusters	Genes in clusters	Genes out of clusters	% genes in clusters
GPCRs*	1,425 (1,079)*	124	668	411	62%	72	904	175	84%
IR	33	3	11	22	33%	6	22	11	67%
DEG-ENaC	12	0	0	12	0%	2	4	8	33%
NPC2	19	4	11	8	58%	4	11	8	58%
SNMP	4	0	0	4	-	0	0	4	-
Lipocalins	10	0	0	10	-	1	2	8	-
GR	7	1	2	5	-	1	3	4	-
Total	1,510	132	692	472	-	86	967	187	-

* In parenthesis the number of complete sequences (Scp), the values used to define gene clusters.

contained proportionally more chemosensory-related genes (Figure S2, Tables S4 and S5).

We operationally used two different *g* values in each species (100−150 kb, and 10−100 kb in *L. ventricosus* and *P. canaliculata*, respectively) to define C_L (Fig. 2A and B). As in Escuer et al. (2022), we choose these values based on the probability of finding two or more chemosensory genes clustered by chance (*p*-value lower than 0.01). For the analysis, we used the number of genes identified in the GPCR family. These *g* values are around the average gene density for each genome (one gene every 110.3 kb, and 20.4 kb for *L. ventricosus* and *P. canaliculata* respectively; Table 1). Although the *p*-value for 100 kb in *P. canaliculata* was 0.042, we included this shared *g* value for comparison purposes.

We found that for both species, the chemosensory-related gene families were mainly organized into clusters (Figures S3 and S4). In *L. ventricosus* and using the $g = 100$ kb, we defined a total of 235 clusters encompassing 989 genes (62% of complete sequences); with $g = 150$ kb, we identified 243 clusters including 1,127 genes (70.3% of the genes present in clusters; Tables 2 and S6; Figure S3). The same trend was observed in the *P. canaliculata* genome. Using the most stringent criteria ($g = 10$ kb), we were able to identify 124 clusters encompassing 668 genes (61.9% of the full length members), a percentage that increased to 83.8% using the more relaxed criterium ($g = 100$ kb; Tables 2, and S7). A similar clustering pattern was observed for both species in the IR/iGluR, DEG-ENaC, and NPC2 gene families (Tables 2, S6–S8).

3.3. Phylogenetic relationships of the chemosensory gene family members

Our phylogenetic analysis using the information from complete GPCR sequences (*Scp* = 3,427, across the three gastropod genomes) revealed that most belong to the Rhodopsin-like (class A) family, as classified under the GRAFS system (Alexander et al., 2021). We identified two major groups in *L. ventricosus* and *P. canaliculata*, one with most members having the 7TM_GPCR_Srw domain, denoted as Serpentine Srw (PF10324) group, and another more diverse group showing the 7TM_1 domain, denoted as Rhodopsin (PF00001) group (Fig. 3). Noticeably, we found that many large clades (with more than 20 sequences) were from the same species (species-specific); twelve in *L. ventricosus* and nine in *P. canaliculata*. For instance, the clade L4 included 354 *L. ventricosus* members, and the P1 of *P. canaliculata,* 329 (Table S9). Since the analysis included only complete proteins located in chromosomes (excluding information from minor scaffolds), the actual number could be greater. We also identified three other clades named H1-3 that encompassed members of the three gastropod species (Fig. 3).

We found that most GPCR genes of *A. californica* conformed to a monophyletic group, sister to the other two gastropod clades (A1 in *Aplysia* group; Fig. 3). Indeed, all the 90 chemosensory-expressed genes identified in Cummins et al. (2009) are in this group, 35 and 51 of them have the 7TM_1 and 7TM_Srw domains, respectively. Interestingly, clade A2 was placed deep within the phylogeny as sister to *L. ventricosus* and *P. canaliculata* members (Fig. 3). The phylogenetic analysis including representatives from other protostomes (Arthropoda, Nematoda, and Platyhelminthes), along with sequences of the Cnidaria phylum, that diverged before the protostomes/deuterostomes split, and the Amoebozoa outgroup uncovered six main GPCR clades, named from I to VI (Fig. 4). Clades I and II included members of the Serpentine Srw (PF10324) group, with clade I being gastropod-specific, whereas II and VI encompassed members of all protostome groups. Clades III to VI included the Rhodopsin (PF00001) group members; with clades III and V containing large groups of gastropod-specific sequences. Clade IV included mainly nematode (*C. elegans*) sequences.

The gene family trees of the other chemosensory-related proteins uncovered different evolutionary dynamics. For instance, in the IR/iGluR gene family tree, we identified several orthologs between Caenogastropoda and *D. melanogaster* IR, a family with some members known to have chemosensory roles in this and other insect species

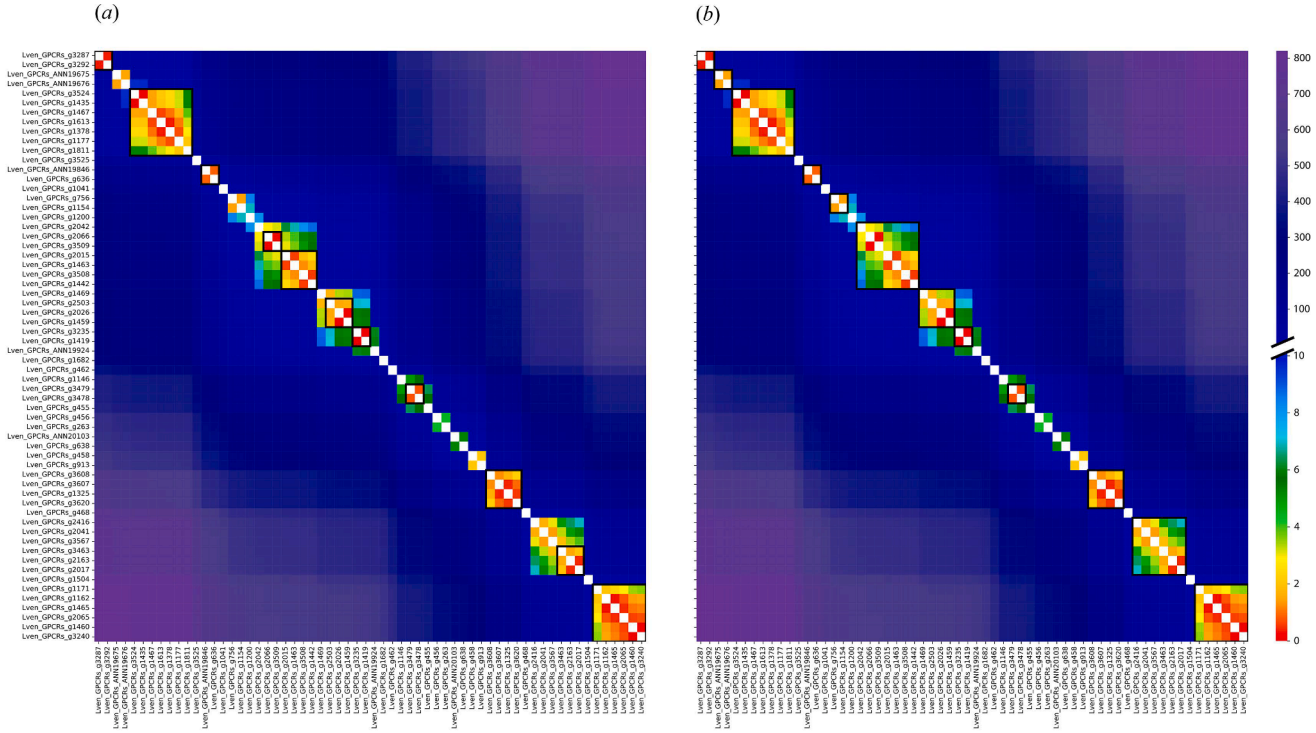

Fig. 2. Genome organization into clusters for the GPCR gene family in chromosome 19 of *L. ventricosus* using a *g* value of 100 kb (panel a) and of 150 kb (panel b). The squares with thicker lines represent the clusters (11 clusters in panel a; 12 clusters in panel b). The range of the colors shows the physical distances on a 100 kb scale.

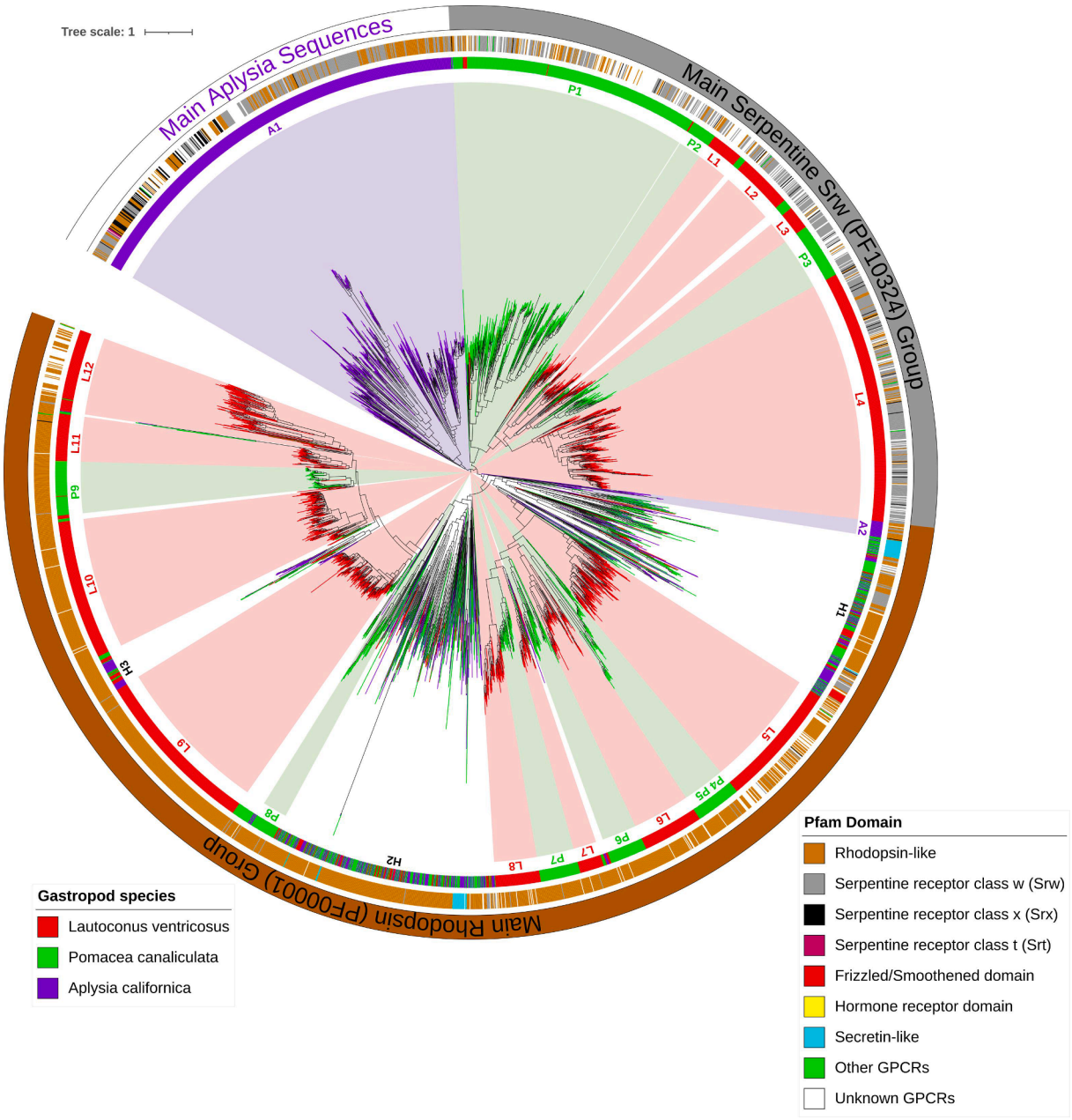

Fig. 3. Phylogenetic relationships among complete GPCR sequences (*Scp*) belonging to the Rhodopsin family (Class A) in *L. ventricosus* (red branches and strip in the inner ring), *P. canaliculata* (green branches and strip in the inner ring), and *A. californica* (purple branches and strip in the inner ring). Three main clades are highlighted in the outer ring: *Aplysia* sequences (purple strip), Serpentine Srw (PF10324) group (gray strip), Rhodopsin (PF0001) group (brown strip). The large gene expansion clades are shadowed in colors. Pfam domains are indicated in the central ring. All clades showed support values greater than 90, except clades H1, P4, P6, H2, P9, L11, and L12 whose values were greater than 63, after 1,000 replicates.The scale bar refers to 1 amino acid substitution per site and the substitution model chosen was JTT + R10. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Figure S5). Despite the absence of large species-specific clades in gastropods, we could identify some gene duplications in *L. ventricosus*. Many of the *D. melanogaster* IRs formed a clade, reflecting either a rapid turnover rate of this family, or a specific diversification in insects. However, we also found old and conserved IRs, such as the putative gastropod homologs of the co-receptor IR25a, and the IR76b, both sour-detecting proteins required in *Drosophila* females for oviposition preference in acid-containing food (Chen and Amrein, 2017). A similar pattern could be observed for the DEG-ENaC gene family (also known as *pkk* genes in *Drosophila*), where the gene tree shows a large clade of *Drosophila* members, clearly separated from those of gastropods, but also a putative homolog in gastropods of the highly conserved receptor *ppk17,* gene involved in sodium ion transmembrane transport and with roles in fluid and salt absorbance, mechanosensation and chemosensation (Zelle et al., 2013) (Figure S6).

3.4. Physical and evolutionary distances

We performed a comparative genomic analysis of pairwise evolutionary distances (measured as the number of amino acid substitutions per site) versus physical distances (in bp) for each GPCR member within and outside genomic clusters. We found, as in Escuer et al., 2022, that the evolutionary distances increased with the physical distance on the chromosome 11 of *L. ventricosus* (Fig. 5a), and this phenomenon was observed in all other chromosomes (Fig. 5b).

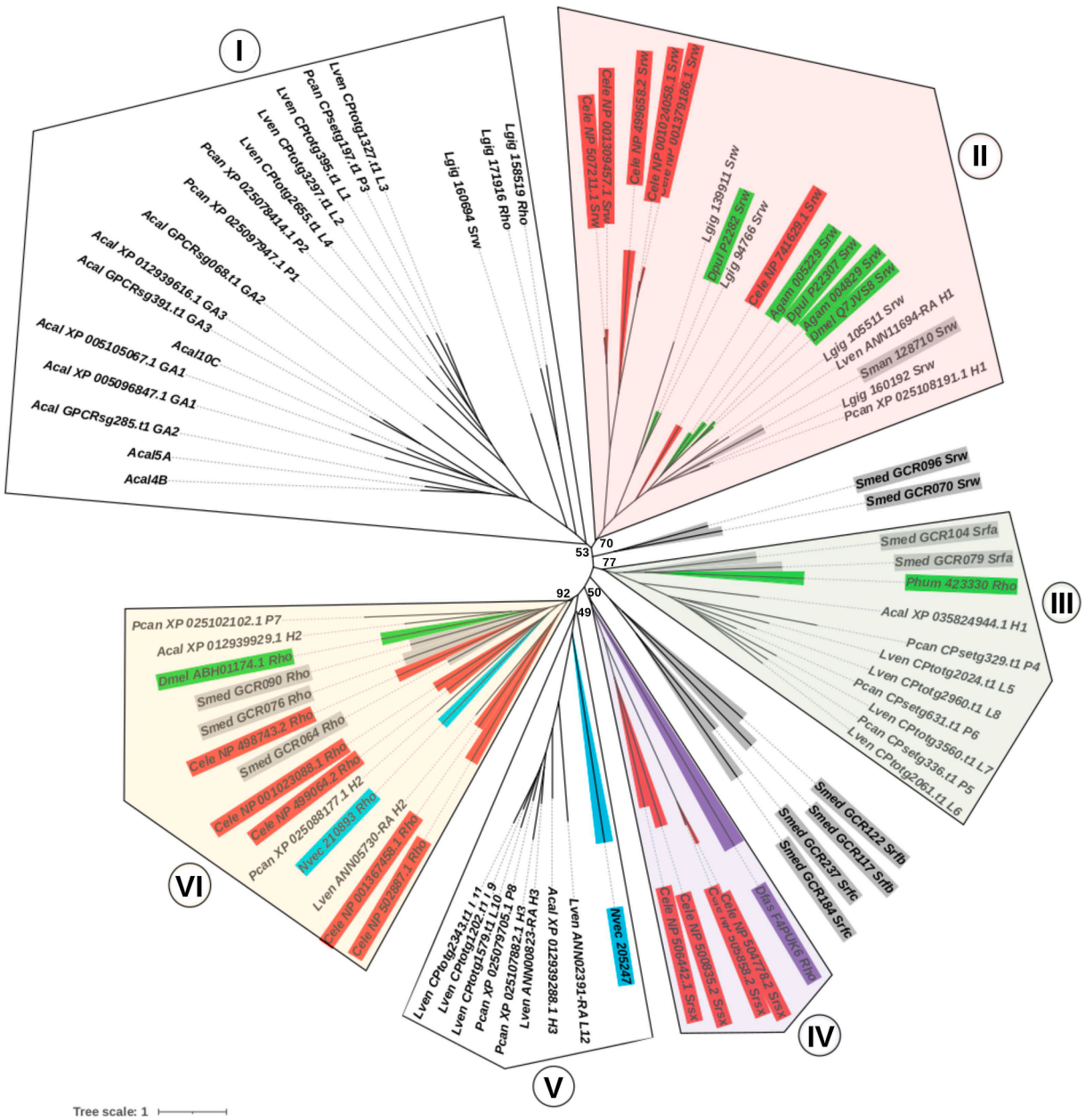

Fig. 4. Phylogenetic relationships among complete GPCR sequences Class A from representatives of different phyla. Mollusk sequences (white background) including representatives of some expansion clades of *L. ventricosus* (Lven), *P. canaliculata* (Pcan), and *A. californica* (Acal), and sequences from the gastropod *Lottia gigantea* (Lgig), the nematode (red background) *C. elegans* (Cele), arthropods (green background) *Anopheles gambiae* (Agam), *Daphnia pulex* (Dpul), *Drosophila melanogaster* (Dmel) and *Pediculus humanus* (Phum), the platyhelminthes (gray background) *Schmidtea mediterranea* (Smed) and *Schistosoma mansoni* (Sman), the cnidarian (cyan background) *Nematostella vectensis* (Nvec), and the amoeba (purple background) *Dyctiostelium fasciculatum* (Dfas). In each clade, the support value after 1,000 replicates is shown. The scale bar refers to 1 amino acid substitution per site and the substitution model chosen was LG + F + R5. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

4.1. Repertoires and clustering of chemosensory-related gene families in gastropods

To our knowledge, there is not a comprehensive report of chemosensory-related gene families in mollusks. Pioneer studies in gastropods, the largest taxonomic class of mollusks, either provided partial information of the proteins expressed in sensory organs (Cummins et al., 2009), or searched for putative orthologs of the main CS families of arthropods on the first fragmented genome drafts of species such as the sea hare *A. californica* (sea slug) and the owl limpet *L. gigantea* (Croset et al., 2010; Krishnan et al., 2014; Saina et al., 2015). Here, we analyzed chromosome-level assemblies and found that the true repertoire of chemosensory-related genes in gastropods is undoubtedly much higher than previously reported, especially in the case of GPCRs. For instance, we found at least 872 GPCR genes in the genome of *A. californica* compared to the 90 receptors previously identified in a transcriptomic study (Cummins et al., 2009). In *L. ventricosus* and *P. canaliculata,* we identified up to 3,022 and 1,425 GPCRs, respectively. To overcome putative genome annotation problems (caused by bioinformatic sensitivity or the highly repetitive nature of the studied genomes), for some analysis we used conservative estimates of gene family size, *Smin* (minimum number of functional sequences) or *Scp* (the

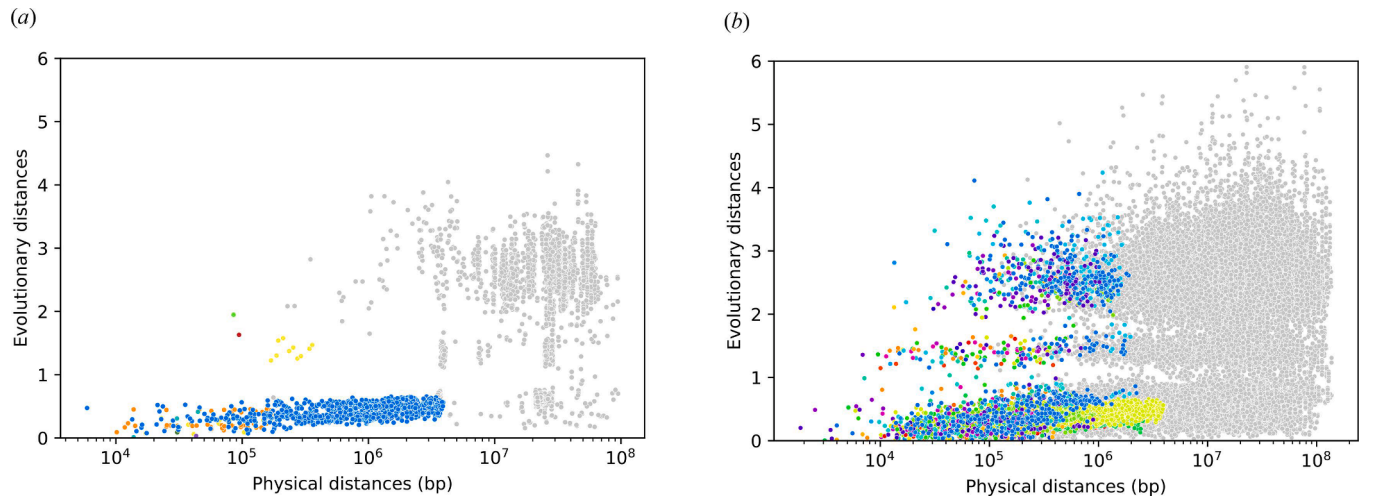

Fig. 5. Plots comparing evolutionary (amino acid substitution per site) and physical (in base pairs) distances between GPCRs copies in *L. ventricosus*. Each dot represents a comparison of two GPCR members of the same chromosome. Colored and gray dots show comparisons within and outside genomic clusters, respectively ($g = 100$ kb). Panel (a) data from chromosome 11; each cluster is depicted in a different color. Panel (b) data from the total data (average for the 35 chromosomes); each chromosome is depicted in a different color.

number of complete members), as in Vizueta et al., 2018. The estimated *Smin* was 2,053 and 1,222 in *L. ventricosus* and *P. canaliculata*, respectively. Here, we found that gastropods exhibit a much higher chemosensory diversity than that exhibited by arthropods, whose chemoreceptor repertoires are usually in the hundreds (Vizueta et al., 2020a), with the exception of some chelicerates; in this latter group, nevertheless the main expanded family encodes the GR (putative gustatory) receptors (Vizueta et al., 2018). In fact, we also identified homologs to these GRs and to the IR/iGluR family, which are well known arthropod chemoreceptors, confirming that the emergence of these families trace back at least to the ancestor of protostomes (Krishnan et al., 2014; Saina et al., 2015). In contrast, we did not find any homolog of the insect OR family, which likely has a more recent origin in arthropods (Robertson, 2015). In summary, gastropods have the largest CS (GPCR) family described thus far in invertebrates, with repertoire sizes similar to those found in many vertebrates (Shi and Zhang, 2009). Nevertheless, it remains necessary to validate that all these identified members really perform a chemosensory function. Indeed, it is expected that not all GPCR genes have a chemosensory role, for instance, while in humans about half of the GPCR members are chemosensory (Di Pizio et al., 2019), this fraction increases to the 88% in *C. elegans* (Vidal et al., 2018). Therefore, we could expect that at least half of the genes identified for this family in gastropods are linked to chemosensory functions, in which case our conclusions remain unchanged. Overall, all these observations could be well accommodated by the Birth-and-Death model of gene family evolution (Nei and Rooney, 2005; Vieira et al., 2007). Unfortunately, the availability of very few highly continuous genomes across gastropods prevents obtaining sound estimations of the turnover rates for the chemosensory families, and to perform a comparative analysis with that of other invertebrates and vertebrates.

We found that the GPCR genes were not randomly distributed within chromosomes; indeed, in *L. ventricosus* and *P. canaliculata* more than 60% of members were arranged into genome clusters. Furthermore, most of these tandemly arranged paralogs are also clustered together in the phylogenetic tree, suggesting recent common ancestry. Remarkably, this uneven distribution of GPCR copies across chromosomes is not associated with the size, or with the number of genes in the chromosomes. Indeed, we found large species-specific clades enriched in genes located on a few chromosomes. We also found that genes within clusters (physically close) exhibit lower genetic distances than those outside them (Fig. 5). Thus, new copies would originate in tandem by unequal crossing-over, they would remain physically close for a while and then progressively move to other genomic regions through mutational mechanisms underlying the formation of structural variations, such as DNA recombination-, replication- and repair-associated processes (Carvalho and Lupski, 2016) or they could be lost by deletions or be retained transiently as pseudogenes. The presence of large species-specific clades indicates that in the gastropod genomes many GPCR members have a recent origin (at least with respect to separation from the other invertebrate groups). This notable number of members, which represents a significant fraction of protein-coding genes encoded in the genome (Table 1), highlights the biological importance (i.e., interaction with the environment, conspecifics, prey, and predators) of the expansion of this chemosensory-related gene family.

The evolutionary relationships of GPCR members are still unclear, especially in protostomes (Nordström et al., 2011). Indeed, the sequence similarity and the type of ligands recognized by these receptors have been the key features for GPCR's classification into five major groups: Glutamate, Rhodopsin, Adhesion, Frizzled, and Secretin under the GRAFS system. Other nomenclature classifies GPCRs into six classes or clans: Rhodopsin-like (Class A), Secretin-like GPCRs (Class B), metabotropic glutamate receptor (Class C), fungal mating pheromone receptors (Class D), cAMP receptors (Class E), and frizzled/smoothened (Class F) (Alexander et al., 2021). Here, we found that most of the GPCR identified in gastropods belong to the Rhodopsin (Class A) family, including members of many of the described subfamilies (Nordström et al., 2011; Saberi et al., 2016), with half of them likely belonging to the *Srw* subfamily.

Remarkably, and despite the large number of specie-specific expansions, we found three clades (H1, H2, and H3) that included members of the three gastropod species with a large number of sequences (e.g., 319 members in H2); likely these members could have a (general) important role in gastropod chemosensation (Fig. 3). The phylogenetic analysis including representatives from other invertebrates (Arthropoda, Nematoda, Platyhelminthes, and Cnidaria), which allowed us to explore the evolution of GPCRs on a larger taxonomic scale (Fig. 4), uncovered both gastropod-specific clades (e.g., clade I), as well as others with members from all analyzed protostome genomes (such as clades II and IV).

4.2. The evolutionary significance of the WGD in the chemosensory-related gene repertoire of caenogastropods

Although successful WGDs can increase the adaptive potential of an organism by driving morphological and physiological innovations, they

are rarely observed in the evolutionary history of animals because they can cause many genetic incompatibilities and meiotic problems (Orr, 1990). In mollusks, there have been reported three putative WGD events, one having occurred in the ancestor of the Stylommatophora (infraclass Heterobranchia), the other close to the ancestor of Neogastropoda (infraclass Caenogastropoda), and the last one within Cephalopoda, concomitant with radiation events of the three groups (Ponder et al., 2008; Hallinan & Lindberg, 2011). The two gastropod ancient WGDs have been supported by comparative analyses involving the chromosome-level genomes of *L. ventricosus* (Pardos-Blas et al., 2021) and *Achatina immaculata* (Liu et al., 2021), which represent Neogastropoda and Stylommatophora, respectively. In the case of *L. ventricosus*, synteny analyses showed that most chromosomal regions of *P. canaliculata* had homolog counterparts in at least two chromosomes in *L. ventricosus*, and the comparison of *hox* gene clusters also agreed with a WGD event (Pardos-Blas et al., 2021). Here, we found that CS family sizes could also reflect the WGD event that occurred in Caenogastropoda, coupled with high gene retention (Table 2). Indeed, we found that the family sizes were nearly double in *L. ventricosus* compared to *P. canaliculata*: GPCR 2.1 times higher; IR/iGluR, 1.7 times; and DEG-ENaC, 1.5 times. The WGD, probably took place approximately between 203 and 155 Mya (Hallinan & Lindberg, 2011), with the elapsed time large enough to have eroded the molecular fingerprint in (evolutionary fast) chemosensory genes via dispensable gene loss (Brunet et al., 2006). If that is the case, the putative adaptive significance of some chemosensory genes would explain the large family sizes. However, although the expansion of these gene families could be decisive in the functional and ecological diversification of Neogastropoda, for instance associated with their active predatory behavior, the WGD could be also concomitant with many non-adaptive gene repertoire changes pre- and post-WGD (Nei et al., 2008). Hence, a more exhaustive and statistically prone analysis of turnover rates and gene copy sequence evolution is required to distinguish between all potential WGD/non-WGD and adaptive/non-adaptive scenarios to understand the biological meaning of the current high diversity of chemosensory-related genes in *L. ventricosus*. In this regard, the future availability of genomes with a wider taxonomic representation of gastropods is still needed to fully understand the evolution of CS genes and the role of WGDs in their diversification.

4.3. Functional significance of candidate genes

Currently, there is little knowledge about how the signal detection and processing occurs in the sensory organs of gastropods, and how these signals are translated into responses at the physiological or behavioral level. Studies in *A. californica* and in the giant triton *Charonia tritonis* (Caenogastropoda: Ranellidae) suggest that chemosensation occurs in cephalic sensory organs, such as the rhinophore and oral tentacles (the proboscis in cone snails), where the expression of GPCRs has been detected (Cummins et al., 2009; Lindberg & Sigwart, 2015; Motti et al., 2022). Another potential target is the osphradium, an organ located in the mantle cavity that is highly developed in predatory species such as cone snails, and also allows the detection of prey odorants through chemoreceptors (Taylor & Miller, 1989).

Our results, uncovering about 1,000–3,000 GPCR genes in each genome of *L. ventricosus, P. canaliculata and A. californica,* highlight the physiological relevance of this family in the studied organisms. Nevertheless, some of the identified GPCR members may likely play a non-chemosensory role (Kang and Koo, 2012), so some extra functional and behavioral experiments, such as deep transcriptomics or knockout of candidate genes, are definitely needed to determine the specific members involved in chemoreception. Similarly, although the chemosensory function of the candidates belonging to other chemosensory gene families (e.g., IR, DEG-ENaC, NPC2, SNMP, and lipocalins) cannot be ruled out, the fact that we detected a relatively reduced number of members of these families preclude any conclusion about its overall chemosensory significance. While they may not likely be primarily responsible for recognising the plethora of chemical cues that a gastropod has to deal with, some of these molecules fulfill highly specific functions.

There is also little knowledge of the role (if any) of soluble proteins, such as NPC2, OBPs, and CSPs in the recognition of target molecules outside insects (Leal, 2013). If they are an adaptation to terrestrial life, as has been suggested (Mollo et al., 2017), marine gastropods, such as *L. ventricosus*, might not need some of these proteins, or may depend on them to a lesser degree, hypothesis that could be tested using genomic data from the recently published genome of the lung slug *Arion vulgaris* (Chen et al., 2022). In marine gastropods, the flow of water through the osphradium (close to the gills), allows the transport of molecules and sediments through the membrane epithelium where they would presumably be detected by GPCRs. This feature could explain why we only detected seven NPC2 genes in *L. ventricosus* (Table 2). On the contrary, despite having lower genome sizes, *P. canaliculata* and *A. californica* encoded more putative chemosensory soluble proteins, 31 and 29 members, respectively, including NPC2 and lipocalin-like OBPs (Tables 2 and S1). Although the three studied gastropod species have an aquatic lifestyle, the lower number of chemosensory soluble proteins in the cone snail, compared to the sea slug and the golden apple snail, might suggest more exposure to air in the latter two, and therefore a greater necessity of these soluble proteins to facilitate the activation of the chemosensory receptors. In fact, the golden apple snail has a lung that allows this species to survive in poorly oxygenated waters, to estivate buried in the mud, and to leave the water for oviposition; suggesting that it could be almost an obligate air-breather (Rodriguez et al., 2021).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Ministerio de Ciencia e Innovación of Spain (MCIN/AEI/10.13039/501100011033; grants PID2019-103947GB-C21 and PID2022-138477NB-C22 to J.R., PID2019-103947GB-C22 and PID2022-138477NB-C21 to R.Z.; FPIs fellowships BES-2017-081195 to J.R.P-B and PRE2020-095592 to V.A.P), and from Comissió Interdepartamental de Recerca I Innovació Tecnològica of Catalonia, Spain (2021SGR00279).

Data accessibility and benefit-sharing

The sequences, alignments, distance matrices, phylogenetic trees, annotation files, and additional information are available in the supplementary data (https://data.mendeley.com/datasets/b9zyx26rdp/2).

Authors' contributions

J.R. and R.Z. conceived the project. J.J.R. and V.P. perform the analyses. J.J.R. draft the first version of the manuscript. J.J.R., V.P., J.R.P-B., A.S.-G., R.Z. and J.R. interpreted the data. All authors revised and approved the final manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2023.107986.

References

Abalde, S., Tenorio, M.J., Afonso, C.M.L., Zardoya, R., 2020. Comparative transcriptomics of the venoms of continental and insular radiations of West African cones. Proc. Biol. Sci./R. Soc. 287 (1929), 20200794. https://doi.org/10.1098/rspb.2020.0794.

Adler, E., Hoon, M.A., Mueller, K.L., Chandrashekar, J., Ryba, N.J., Zuker, C.S., 2000. A novel family of mammalian taste receptors. Cell 100 (6), 693–702. https://doi.org/10.1016/s0092-8674(00)80705-9.

Albano, P.G., 2021. Biology and evolution of the MolluscaW. F.PonderD. R. LindberghBoca Raton, FL: CRC Press, Taylor & Francis Group, 2019–2020. Two volumes: xxiii-900 + xx-870 pp. ISBN: 9780815361695 (volume 1), 9780815361848 (volume 2). Hardcover: 190 £ + 190 £. *Marine Ecology* , *42*(4). https://doi.org/10.1111/maec.12645.

Alexander, S.P., Kelly, E., Mathie, A., Peters, J.A., Veale, E.L., Armstrong, J.F., Faccenda, E., Harding, S.D., Pawson, A.J., Southan, C., Buneman, O.P., Cidlowski, J.A., Christopoulos, A., Davenport, A.P., Fabbro, D., Spedding, M., Striessnig, J., Davies, J. A., Ahlers-Dannen, K.E., et al., 2021. The concise guide to pharmacology 2021/22: Introduction and Other Protein Targets. Brit. J. Pharmacol. 178 (Suppl 1), S1–S26. https://doi.org/10.1111/bph.15537.

Angeli, S., Ceron, F., Scaloni, A., Monti, M., Monteforti, G., Minnocci, A., Petacchi, R., Pelosi, P., 1999. Purification, structural characterization, cloning and immunocytochemical localization of chemoreception proteins from *Schistocerca gregaria*. Euro. J. Biochem./FEBS 262 (3), 745–754. https://doi.org/10.1046/j.1432-1327.1999.00438.x.

Benton, R., Dessimoz, C., Moi, D., 2020. A putative origin of the insect chemosensory receptor superfamily in the last common eukaryotic ancestor. eLife 9. https://doi.org/10.7554/eLife.62507.

Böhme, I., Beck-Sickinger, A.G., 2009. Illuminating the life of GPCRs. Cell Commun. Signal 7, 16. https://doi.org/10.1186/1478-811X-7-16.

Brunet, F.G., Roest Crollius, H., Paris, M., Aury, J.-M., Gibert, P., Jaillon, O., Laudet, V., Robinson-Rechavi, M., 2006. Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. Mol. Biol. Evol. 23 (9), 1808–1816. https://doi.org/10.1093/molbev/msl049.

Buck, L., Axel, R., 1991. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell 65 (1), 175–187. https://doi.org/10.1016/0092-8674(91)90418-x.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinf. 10, 421. https://doi.org/10.1186/1471-2105-10-421.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25 (15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Carvalho, C.M.B., Lupski, J.R., 2016. Mechanisms underlying structural variant formation in genomic disorders. Nat. Rev. Genet. 17 (4), 224–238. https://doi.org/10.1038/nrg.2015.25.

Chen, Y., Amrein, H., 2017. Ionotropic receptors mediate *drosophila* oviposition preference through sour gustatory receptor neurons. Curr. Biol. 27 (18), 2741–2750. e4. https://doi.org/10.1016/j.cub.2017.08.003.

Chen, Z., Doğan, Ö., Guiglielmoni, N., Guichard, A., Schrödl, M., 2022. Pulmonate slug evolution is reflected in the de novo genome of *Arion vulgaris* Moquin-Tandon, 1855. Sci. Rep. 12, 14226. https://doi.org/10.1038/s41598-022-18099-7.

Clyne, P.J., Warr, C.G., Freeman, M.R., Lessing, D., Kim, J., Carlson, J.R., 1999. A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. Neuron 22 (2), 327–338. https://doi.org/10.1016/s0896-6273(00)81093-4.

Croset, V., Rytz, R., Cummins, S.F., Budd, A., Brawand, D., Kaessmann, H., Gibson, T.J., Benton, R., 2010. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. PLoS Genet 6 (8), e1001064. https://doi.org/10.1371/journal.pgen.1001064.

Cummins, S.F., Erpenbeck, D., Zou, Z., Claudianos, C., Moroz, L.L., Nagle, G.T., Degnan, B.M., 2009. Candidate chemoreceptor subfamilies differentially expressed in the chemosensory organs of the mollusc *Aplysia*. BMC Biol. 7, 28. https://doi.org/10.1186/1741-7007-7-28.

Derby, C.D., Kozma, M.T., Senatore, A., Schmidt, M., 2016. Molecular mechanisms of reception and perireception in crustacean chemoreception: a comparative review. Chem. Senses 41 (5), 381–398. https://doi.org/10.1093/chemse/bjw057.

Di Pizio, A., Behrens, M., Krautwurst, D., 2019. Beyond the flavour: the potential druggability of chemosensory G protein-coupled receptors. Int. J. Mol. Sci. 20 (6), 1402. https://doi.org/10.3390/ijms20061402.

Eddy, S.R., 2011. Accelerated Profile HMM Searches. PLoS Comput. Biol. 7 (10), e1002195 https://doi.org/10.1371/journal.pcbi.1002195.

Escuer, P., Pisarenco, V.A., Fernández-Ruiz, A.A., Vizueta, J., Sánchez-Herrero, J.F., Arnedo, M.A., Sánchez-Gracia, A., Rozas, J., 2022. The chromosome-scale assembly of the Canary Islands endemic spider *Dysdera silvatica* (Arachnida, Araneae) sheds light on the origin and genome structure of chemoreceptor gene families in chelicerates. Mol. Ecol. Resour. 22 (1), 375–390. https://doi.org/10.1111/1755-0998.13471.

Gough, J., Karplus, K., Hughey, R., Chothia, C., 2001. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. J. Mol. Biol. 313 (4), 903–919. https://doi.org/10.1006/jmbi.2001.5080.

Hallinan, N.M., Lindberg, D.R., 2011. Comparative analysis of chromosome counts infers three paleopolyploidies in the Mollusca. Gen. Biol. Evol. 3, 1150–1163. https://doi.org/10.1093/gbe/evr087.

Hayes, K.A., Joshi, R.C., Thiengo, S.C., Cowie, R.H., 2008. Out of South America: multiple origins of non-native apple snails in Asia. Divers. Distrib. 14 (4), 701–712. https://doi.org/10.1111/j.1472-4642.2008.00483.x.

Hilger, D., Masureel, M., Kobilka, B.K., 2018. Structure and dynamics of GPCR signaling complexes. Nat. Struct. Mol. Biol. 25 (1), 4–12. https://doi.org/10.1038/s41594-017-0011-7.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30 (9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Comp. Appl. Biosci.: CABIOS 8 (3), 275–282. https://doi.org/10.1093/bioinformatics/8.3.275.

Kang, N., Koo, J., 2012. Olfactory receptors in non-chemosensory tissues. BMB Rep. 45 (11), 612–622. https://doi.org/10.5483/bmbrep.2012.45.11.232.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780. https://doi.org/10.1093/molbev/mst010.

Knudsen, B., Kohn, A.B., Nahir, B., McFadden, C.S., Moroz, L.L., 2006. Complete DNA sequence of the mitochondrial genome of the sea-slug, *Aplysia californica*: conservation of the gene order in Euthyneura. Mol. Phylogenet. Evol. 38 (2), 459–469. https://doi.org/10.1016/j.ympev.2005.08.017.

Krishnan, A., Almén, M.S., Fredriksson, R., Schiöth, H.B., 2014. Insights into the origin of nematode chemosensory GPCRs: putative orthologs of the Srw family are found across several phyla of protostomes. PLoS One 9 (3), e93048. https://doi.org/10.1371/journal.pone.0093048.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Transmembrane helices predicted using TMHMM: Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580. https://doi.org/10.1006/jmbi.2000.4315.

Kumar, S., Stecher, G., Peterson, D., Tamura, K., 2012. MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. Bioinformatics 28 (20), 2685–2686. https://doi.org/10.1093/bioinformatics/bts507.

Lach, L., Britton, D.K., Rundell, R.J., Cowie, R.H., 2000. Food preference and reproductive plasticity in an invasive freshwater snail. Biol. Invasions 2 (4), 279–288. https://doi.org/10.1023/a:1011461029986.

Leal, W.S., 2013. Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. Annu. Rev. Entomol. 58, 373–391. https://doi.org/10.1146/annurev-ento-120811-153635.

Lemarcis, T., Fedosov, A.E., Kantor, Y.I., Abdelkrim, J., Zaharias, P., Puillandre, N., 2022. Neogastropod (Mollusca, Gastropoda) phylogeny: a step forward with mitogenomes. Zool. Scr. 51 (5), 550–561. https://doi.org/10.1111/zsc.12552.

Letunic, I., Bork, P., 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucl. Acids Res. 49 (W1), W293–W296. https://doi.org/10.1093/nar/gkab301.

Li, X., Staszewski, L., Xu, H., Durick, K., Zoller, M., Adler, E., 2002. Human receptors for sweet and umami taste. Proc. Nat. Acad. Sci. U.S.A. 99(7), 4692–4696. https://doi.org/10.1073/pnas.072090199.

Liberles, S.D., Buck, L.B., 2006. A second class of chemosensory receptors in the olfactory epithelium. Nature 442 (7103), 645–650. https://doi.org/10.1038/nature05066.

Lindberg, D.R., Sigwart, J.D., 2015. What is the molluscan osphradium? A reconsideration of homology. Zoologischer Anzeiger-A J. Comparat. https://www.sciencedirect.com/science/article/pii/S0044523115000303?casa_token=_cbCJYM7WpgAAAAA:V7SmSEwQYW0pukLldBalVBKBKkGM22yyJon5kF1YZ54G1MM5mho1KLynFNfqi4jXi36t7mFFWm6i.

Liu, C., Zhang, Y., Ren, Y., Wang, H., Li, S., Jiang, F., Yin, L., Qiao, X., Zhang, G., Qian, W., Liu, B., Fan, W., 2018. The genome of the golden apple snail *Pomacea canaliculata* provides insight into stress tolerance and invasive adaptation. GigaScience 7 (9). https://doi.org/10.1093/gigascience/giy101.

Liu, C., Ren, Y., Li, Z., Hu, Q., Yin, L., Wang, H., Qiao, X., Zhang, Y., Xing, L., Xi, Y., Jiang, F., Wang, S., Huang, C., Liu, B., Liu, H., Wan, F., Qian, W., Fan, W., 2021. Giant African snail genomes provide insights into molluscan whole-genome duplication and aquatic-terrestrial transition. Mol. Ecol. Resour. 21 (2), 478–494. https://doi.org/10.1111/1755-0998.13261.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37 (5), 1530–1534. https://doi.org/10.1093/molbev/msaa015.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., Bateman, A., 2021. Pfam: The protein families database in 2021. Nucl. Acids Res. 49 (D1), D412–D419. https://doi.org/10.1093/nar/gkaa913.

Mollo, E., Garson, M.J., Polese, G., Amodeo, P., Ghiselin, M.T., 2017. Taste and smell in aquatic and terrestrial environments. Nat. Prod. Rep. 34 (5), 496–513. https://doi.org/10.1039/c7np00008a.

Motti, C.A., Cummins, S.F., Hall, M.R., 2022. A review of the giant triton (*Charonia tritonis*), from exploitation to coral reef protector? Diversity 14 (11), 961. https://doi.org/10.3390/d14110961.

Mueck, K., Deaton, L.E., Lee, A., 2020. Microscopic anatomy of the gill and lung of the apple snail *Pomacea maculata*, with notes on the volume of the lung. J. Shellfish. Res. 39 (1), 125–132. https://doi.org/10.2983/035.039.0112.

Nei, M., Niimura, Y., Nozawa, M., 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. Nat. Rev. Genet. 9 (12), 951–963. https://doi.org/10.1038/nrg2480.

Nei, M., Rooney, A.P., 2005. Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. 39, 121–152. https://doi.org/10.1146/annurev.genet.39.073003.112240.

Nordström, K.J.V., Sällman Almén, M., Edstam, M.M., Fredriksson, R., Schiöth, H.B., 2011. Independent HHsearch, Needleman–Wunsch-Based, and motif analyses reveal the overall hierarchy for most of the G protein-coupled receptor families. Mol. Biol. Evol. 28 (9), 2471–2480. https://doi.org/10.1093/molbev/msr061.

Orr, H.A., 1990. "why polyploidy is rarer in animals than in plants" revisited. Am. Nat. 136 (6), 759–770. https://doi.org/10.1086/285130.

Osca, D., Templado, J., Zardoya, R., 2015. Caenogastropod mitogenomics. Mol. Phylogenet. Evol. 93, 118–128. https://doi.org/10.1016/j.ympev.2015.07.011.

Pardos-Blas, J.R., Irisarri, I., Abalde, S., Afonso, C.M.L., Tenorio, M.J., Zardoya, R., 2021. The genome of the venomous snail *Lautoconus ventricosus* sheds light on the origin of conotoxin diversity. GigaScience 10 (5). https://doi.org/10.1093/gigascience/giab037.

Pelosi, P., Maida, R., 1995. Odorant-binding proteins in insects. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 111 (3), 503–514. https://doi.org/10.1016/0305-0491(95)00019-5.

Pikielny, C.W., 2012. Sexy DEG/ENaC channels involved in gustatory detection of fruit fly pheromones. Sci. Signal. 5 (249), e48. https://doi.org/10.1126/scisignal.2003555.

Ponder, W.E., Ponder, W., Lindberg, D.R., 2008. Phylogeny and Evolution of the Mollusca. University of California Press. https://play.google.com/store/books/details?id=EzElDQAAQBAJ.

Rawlings, T.A., Hayes, K.A., Cowie, R.H., Collins, T.M., 2007. The identity, distribution, and impacts of non-native apple snails in the continental United States. BMC Evol. Biol. 7, 97. https://doi.org/10.1186/1471-2148-7-97.

Robertson, H.M., 2015. The insect chemoreceptor superfamily is ancient in animals. Chem. Senses 40 (9), 609–614. https://doi.org/10.1093/chemse/bjv046.

Robertson, H.M., 2019. Molecular evolution of the major arthropod chemoreceptor gene families. Annu. Rev. Entomol. 64, 227–242. https://doi.org/10.1146/annurev-ento-020117-043322.

Rodriguez, C., Prieto, G.I., Vega, I.A., Castro-Vazquez, A., 2021. Morphological grounds for the obligate aerial respiration of an aquatic snail: functional and evolutionary perspectives. PeerJ 9, e10763. https://doi.org/10.7717/peerj.10763.

Saberi, A., Jamal, A., Beets, I., Schoofs, L., Newmark, P.A., 2016. GPCRs direct germline development and somatic gonad function in planarians. PLoS Biol 14 (5), e1002457. https://doi.org/10.1371/journal.pbio.1002457.

Saina, M., Busengdal, H., Sinigaglia, C., Petrone, L., Oliveri, P., Rentzsch, F., Benton, R., 2015. A cnidarian homologue of an insect gustatory receptor functions in developmental body patterning. Nat. Commun. 6, 6243. https://doi.org/10.1038/ncomms7243.

Sánchez-Gracia, A., Vieira, F.G., Rozas, J., 2009. Molecular evolution of the major chemosensory gene families in insects. Heredity 103 (3), 208–216. https://doi.org/10.1038/hdy.2009.55.

Sayers, E.W., Cavanaugh, M., Clark, K., Pruitt, K.D., Schoch, C.L., Sherry, S.T., Karsch-Mizrachi, I., 2021. GenBank. Nucl. Acids Res. 49 (D1), D92–D96. https://doi.org/10.1093/nar/gkaa1023.

Shi, P., Zhang, J., 2009. Extraordinary diversity of chemosensory receptor gene repertoires among vertebrates. Results Probl. Cell Differ. 47, 1–23. https://doi.org/10.1007/400_2008_4.

Storch, J., Xu, Z., 2009. Niemann-Pick C2 (NPC2) and intracellular cholesterol trafficking. BBA 1791 (7), 671–678. https://doi.org/10.1016/j.bbalip.2009.02.001.

Taylor, J.D., Miller, J.A., 1989. The morphology of the osphradium in relation to feeding habits in meso- and neogastropods. J. Molluscan Stud. 55 (2), 227–237. https://doi.org/10.1093/mollus/55.2.227.

Uribe, J.E., Irisarri, I., Templado, J., Zardoya, R., 2019. New patellogastropod mitogenomes help counteracting long-branch attraction in the deep phylogeny of gastropod mollusks. Mol. Phylogenet. Evol. 133, 12–23. https://doi.org/10.1016/j.ympev.2018.12.019.

Vidal, B., Aghayeva, U., Sun, H., Wang, C., Glenwinkel, L., Bayer, E.A., Hobert, O., 2018. An atlas of *Caenorhabditis elegans* chemoreceptor expression. PLoS Biol 16 (1), e2004218. https://doi.org/10.1371/journal.pbio.2004218.

Vieira, F.G., Rozas, J., 2011. Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. Genome Biol. Evol. 3, 476–490. https://doi.org/10.1093/gbe/evr033.

Vieira, F.G., Sánchez-Gracia, A., Rozas, J., 2007. Comparative genomic analysis of the odorant-binding protein family in 12 *Drosophila* genomes: purifying selection and birth-and-death evolution. Genome Biol. 8 (11), R235. https://doi.org/10.1186/gb-2007-8-11-r235.

Vizueta, J., Frías-López, C., Macías-Hernández, N., Arnedo, M.A., Sánchez-Gracia, A., Rozas, J., 2017. Evolution of chemosensory gene families in arthropods: insight from the first inclusive comparative transcriptome analysis across spider appendages. Genome Biol. Evol. 9 (1), 178–196. https://doi.org/10.1093/gbe/evw296.

Vizueta, J., Rozas, J., Sánchez-Gracia, A., 2018. Comparative genomics reveals thousands of novel chemosensory genes and massive changes in chemoreceptor repertories across chelicerates. Genome Biol. Evol. 10 (5), 1221–1236. https://doi.org/10.1093/gbe/evy081.

Vizueta, J., Escuer, P., Frías-López, C., Guirao-Rico, S., Hering, L., Mayer, G., Rozas, J., Sánchez-Gracia, A., 2020a. Evolutionary history of major chemosensory gene families across Panarthropoda. Mol. Biol. Evol. 37 (12), 3601–3615. https://doi.org/10.1093/molbev/msaa197.

Vizueta, J., Sánchez-Gracia, A., Rozas, J., 2020b. bitacora: a comprehensive tool for the identification and annotation of gene families in genome assemblies. Mol. Ecol. Resour. 20 (5), 1445–1452. https://doi.org/10.1111/1755-0998.13202.

Zelle, K.M., Lu, B., Pyfrom, S.C., Ben-Shahar, Y., 2013. The genetic architecture of degenerin/epithelial sodium channels in *Drosophila*. G3: Genes, Genomes. Genetics 3 (3), 441–450. https://doi.org/10.1534/g3.112.005272.