The genome of the simultaneously hermaphroditic snail Lymnaea stagnalis reveals an evolutionary expansion of FMRFamide-like receptors

Joris M Koene, Daniel J. Jackson, Yumi Nakadera, Nicolas Cerveau, Mohammed-amin Madoui, Benjamin Noel, Véronique Jamilloux, Julie Poulain, Karine Labadie, Corinne Da Silva, et al.

The genome of the simultaneously hermaphroditic snail Lymnaea stagnalis reveals an evolutionary expansion of FMRFamide-like receptors

Joris M. Koene
Amsterdam Institute for Life and Environment (A-LIFE), Vrije Universiteit
Daniel J. Jackson
Georg-August University of Göttingen
Yumi Nakadera
Amsterdam Institute for Life and Environment (A-LIFE), Vrije Universiteit
Nicolas Cerveau
Georg-August University of Göttingen
Mohammed-Amin Madoui
SEPIA, Institut François Jacob, Université Paris Saclay
Benjamin Noel
Génomique Métabolique, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay
Véronique Jamilloux
PROSE, INRAE, Paris-Saclay University
Julie Poulain
Génomique Métabolique, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay
Karine Labadie
Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay
Corinne Da Silva
Génomique Métabolique, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay
Angus Davison
University of Nottingham
Zhong-Ping Feng
King's College
Coen M. Adema
University of New Mexico
Christophe Klopp
INRAE, Applied Mathematics and Informatics of Toulouse, MIAT UR875
Jean-Marc Aury
Génomique Métabolique, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay
Patrick Wincker

Génomique Métabolique, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay

Marie-Agnès Coutellec (marie-agnes.coutellec@inrae.fr)

DECOD (Ecosystem Dynamics and Sustainability), INRAE, L'Institut Agro

---

Article

Keywords: Molluscan genome, nuclear receptors, GPCRs, gene expansion, simultaneous hermaphroditism

Posted Date: February 12th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3948809/v1

Additional Declarations: No competing interests reported.

Abstract

The great pond snail *Lymnaea stagnalis* has served as a model organism for over a century in diverse disciplines such as neurophysiology, evolution, ecotoxicology and developmental biology. To support both established uses and newly emerging research interests we have performed whole genome sequencing (~ 176 x depth), assembly and annotation of a single individual derived from an inbred line. These efforts resulted in a final assembly of 943 Mb (L50 = 257; N50 = 957,215) with a total of 22,499 predicted gene models. As a first step towards understanding the hermaphroditic reproductive biology of *L. stagnalis*, we identified molecular receptors, specifically nuclear receptors (including newly discovered 2xDNA binding domain-NRs), G protein-coupled receptors, and receptor tyrosine kinases, that may be involved in the cellular specification and maintenance of simultaneously active male and female reproductive systems. A phylogenetic analysis of one particular family of GPCRs (Rhodopsin neuropeptide FMRFamide-receptor-like genes) shows a remarkable expansion that coincides with the occurrence of simultaneous hermaphroditism in the Euthyneura gastropods. As some GPCRs and NRs also showed qualitative differences in expression in female (albumen gland) and male (prostate gland) organs, it is possible that separate regulation of male and female reproductive processes may in part have been enabled by an increased abundance of receptors in the transition from a separate-sexed state to a hermaphroditic condition. These findings will support efforts to pair receptors with their activating ligands, and more generally stimulate deeper insight into the mechanisms that underlie the modes of action of compounds involved in neuroendocrine regulation of reproduction, induced toxicity, and development in *L. stagnalis*, and molluscs in general.

Introduction

The ability of animals to perceive and react to their environments is key to their survival and evolutionary fitness. The study of such sensory abilities across the metazoan tree of life has led to deep insights across many fields and has deepened our understanding of how the abilities to taste, hear, touch, see and smell evolved [1–4]. Similarly, highly sensitive and consistent cellular responses to a local environment are prerequisites for the operation of robust developmental and reproductive programs. With the growing library of metazoan genome sequences, it is clear that a rich diversity of molecular receptors underlies many aspects of these sensory abilities [5–7]. For example, the family of receptors known as G protein-coupled receptors (GPCRs) is one of the largest protein superfamilies in the human genome [8]. However, our general understanding of the ligands that bind these receptors in most cases is lacking, and concerted efforts to pair these receptors with their activating ligands (increasingly with new technologies such as artificial intelligence), are underway [9–12]. Given a history across a variety of disciplines (from behavioural, to neural and molecular - reviewed in [13]), the great pond snail *L. stagnalis* has the potential to make significant contributions to these efforts. Many of these previous investigations were focused on basic biological processes such as respiration, heart rate, egg laying, mating, feeding and more, and have also resulted in *Lymnaea* becoming a standard species for ecotoxicology studies (OECD TG243 [14]).

More recently, knowledge of fundamental neuro-biological processes has provided important insight into learning and memory in *L. stagnalis*, and how these processes are disturbed by human and/or environmental factors (e.g. [15, 16, 17]). Likewise, ecotoxicological compounds and immune elicitors in the environment have been shown to affect this species in terms of behavioural, developmental, endocrine, immunological and oxidative stress responses (e.g., [18, 19–24]). Importantly, some of these responses have been linked to molecular profiles in the central nervous system (e.g. [25, 26–28]). For example, recent work has highlighted the role of the neurotransmitter serotonin during development in the egg (e.g. [29, 30]), for which the extensively documented ontology development of this species has proven valuable (e.g. [31, 32–35]). *Lymnaea* has also proven to be a valuable model for investigations focused on the cellular and molecular mechanisms of memory [36–38], biomineralisation [39–41], and has been instrumental in elucidating the key genes involved in body-plan chirality, thanks to a naturally occurring sinistral mutant line that could be back-crossed, and experimentally and transcriptionally investigated alongside the normal, dextral type [42–45]. A similar approach can be used to identify the gene that is responsible for the recently discovered shell colour variant, referred to as Ginger [46].

A fascinating feature of *L. stagnalis* that has received increasing attention over the past decades is its reproduction. As a simultaneous hermaphrodite, these snails can mate in the male and female role, but within a single mating interaction one individual must choose which role to perform; roles can subsequently be immediately swapped with the same or a different partner [47]. The particulars of this behaviour allow for experimentally disentangling male and female processes and underlying motivations [48–51], something that is difficult to do in many other hermaphroditic species [52, 53]. As a result, the great pond snail has played a central role in driving evolutionary research on hermaphroditic reproductive processes including selfing, mate choice, sexual conflict, and sperm competition [54–59].

To advance these fields, the *Lymnaea* research community requires an annotated genome sequence to facilitate the use of contemporary sequencing methods (RNA-Seq, single cell sequencing, ATACseq), proteomics, comparative genomics and functional genomics (e.g., RNAi and CRISPR-Cas9). Moreover, such a resource will be the foundation for the identification and characterisation of novel molecular receptors. The process of ligand/receptor de-orphanisation will enable the study of their role in basic biological processes and inform the ecological risk assessment of chemicals such as suspected endocrine disrupters (see the OECD framework for endocrine disruption; [60, 61]). In addition, such an inventory will provide a framework to investigate whether this hermaphrodite (and others) deviates from separate-sexed species in its use and diversity of receptors, and how such life history strategies may have evolved.

To promote further progress, we sequenced, assembled and annotated the genome of *L. stagnalis*, and searched for gene family expansions using orthology analyses. We focused on identifying and analysing nuclear receptors (NRs), G-protein coupled receptors (GPCRs), and receptor tyrosine kinases (RTKs). These three highly conserved superfamilies were selected based on their pivotal roles in endocrine systems (NRs: [62]), signalling pathways and reproduction (GPCRs: [63]); (RTKs: [64, 65]), and possible

crosstalk among the signalling pathways that are activated by these different types of receptors [66]. For a subset of these receptors, we performed *in situ* hybridisation (ISH) on early developmental stages and histological sections of sex-specific organs of adult *L. stagnalis* to localise their expression. Our annotation effort complements a previous effort devoted to ion-channels and ionotropic receptors in the central nervous system [67]. Here we focus on the largest family of receptors, GPCRs, to assess the difference between simultaneously hermaphroditic gastropod molluscs (such as *L. stagnalis*) and gastropods that manifest one sex at a time. Given that GPCRs are involved in numerous behaviours across the Metazoa including copulation and egg-laying [68], we anticipate that our approach will contribute to evolutionary hypotheses underlying the evolution and diversification of receptors and their roles in maintaining the simultaneous hermaphroditic state.

Results

Genome sequencing and assembly

The *Lymnaea stagnalis* genome was sequenced using a combination of both Illumina paired-end and mate-pair reads. The resulting draft assembly spans 943 Mb and comprises 6,640 scaffolds with lengths ranging from 2,000 to 7,106,496 bp, as outlined in Table 1. This assembly is 250 Mb lower than the previously measured genome size of 1.193 Gb (C-value = 1.22 pg [69]). This discrepancy is consistent with the general trend observed in eukaryotes (see [70]).

Table 1
Summary of the *L. stagnalis* genome assembly.

Metric	Final assembly Contigs	Final assembly Scaffolds
Assembly size	873,879,680	942,996,421
Number	64,200	6,640
N50	35,351	957,215
L50	6,921	297
N90	7,115	151,064
L90	27,651	1,178
Largest	315,608	7,106,496
GC%		37.55%
%N's		7.33%

Repeat content

Repetitive elements spanned 37.89% of the 943 Mb-genome. Using a semi-automated process (described in the materials and methods section), we annotated the *L. stagnalis* genome with 2,643 transposable element (TE) consensus sequences (provisionally available via INRAE data repository:

(https://entrepot.recherche.data.gouv.fr/privateurl.xhtml?token=1ad8e7c4-fa43-4002-b99f-c236eae7ba01), giving 883,147 TE copies (24,190 of which are full-length copies (FLCs). In the consensus library, 1,517 TEs were classified as retrotransposons (Class I) and 774 as DNA transposons (Class II), 243 were not classified and 105 as Potential Host Genes (PHGs). Most of these consensus sequences (2,604, i.e., 99%) have one or more full-length copy, indicating that a consensus sequence is likely a reasonable estimate of the common ancestor of any one TE. The resulting annotation was used to mask the TE landscape in the genome.

Annotation

Following repeat masking of the genome, we employed *ab initio*, transcriptome- and homology-based methods to predict a set of 22,499 protein-coding genes (see Table 2 for detailed gene characteristics). This number is somewhat lower than in other annotated gastropod genomes (e.g., 23,851 genes in *L. gigantea*, 28,683 in *A. californica*, 36,943 in *B. glabrata*), and lower than the average number of genes found in a panel of 20 representative Metazoa (see OrthoFinder analysis).

Table 2
Characteristics of the L. stagnalis genome annotation.

Parameter	value
Assembly size	943 Mbp
# predicted genes	22,499
# predicted genes without intron	3,492 (15.5%)
# exon per gene avg:med	7.59:5
Gene size (nt) avg:med	15,299:8,502
CDS size (nt) avg:med	1,375:993
Intron size (nt) avg:med	1,611:604
Complete BUSCOs	98.5% (vs, odb10 Eukaryota)
Complete and single-copy BUSCOs	97.3%
Complete and duplicated BUSCOs	1.2%
Fragmented BUSCOs	1.2%
Missing BUSCOs	0.3%
Total BUSCO groups searched	255
*# predicted genes with domain**	19,387 (86%)
# predicted genes with InterPro domain**	16,622 (74%)
# predicted genes with Pfam domain**	15,088 (67%)
**at least one protein domain*	

Mitochondrial genome

BLAST searches returned mitogenomic sequence similarities for the full length of Lsta_scaffold2639. As indicated by a 127 bp sequence overlap at the termini, this scaffold represents a circular sequence. Nanopore reads confirmed that the assembled sequence is contiguous, and gene annotations revealed the complete standard gene complement for a metazoan mitogenome. The 13,834 bp mitogenome of *L. stagnalis* (Genbank MW221941) is 28.21% GC-rich, different from the 37.55% AT content of the nuclear genome. Several protein coding genes are delimited by downstream tRNA genes, causing the ORF to terminate with a truncated stop codon (T- or TA-) that is completed by polyadenylation of the mRNA transcripts. This mitogenome sequence has 94% identity with another, independently characterised *L. stagnalis* mitogenome (13,807 bp, MT874495). The different lengths of these two mitogenomes are mostly due to indels in regions with tRNA- and rRNA genes, a two bp indel occurs at the 5' start of ND4L. These mitogenome sequences align closely but contain abundant single nucleotide variants (including

several non-synonymous substitutions) within protein coding genes. The organisation of the 22 tRNA genes, two rRNA genes plus 13 protein coding genes (Supplementary Figure SF1) of the *L. stagnalis* mitogenome is generally similar to that of other lymnaeid (and hygrophilid) snail species, with a few minor differences in the location of tRNA genes.

Receptor identification

Receptor Tyrosine Kinases

A total of 32 gene predictions (including manually curated versions) were identified as Receptor Tyrosine Kinases (RTKs), distributed across 18 of the 20 families or types (see Supplementary Table ST1). Among these, 12 gene predictions did not contain expected (canonical) domains. We found no gene prediction representing type III or XI RTKs, while two gene predictions could not be assigned to a specific RTK type.

Nuclear receptors

In total, 47 nuclear receptors (NRs) were identified from automatic and manual gene predictions, 39 of which containing the complete sequence of both conserved domains, i.e., the DNA-binding domain (DBD) and ligand binding domain (LBD) (Supplementary Tables ST2 and ST3). All 6 traditional NR families were represented as follows: NR1 (28 predictions), NR2 (12), NR3 (3), NR4 (1), NR5 (2), NR6 (1). However, in terms of NR nomenclature, these corresponded to only 29 different NRs, as several gene predictions were found to have the same annotation (e.g., NR1F1_RORA, represented by four different gene predictions, that share low levels of domain sequence similarity (from 28.9–41.5%).

Significantly, we found two NRs with 2DBDs and one LBD. The occurrence of such an atypical NR structure was first reported in 2006 in the parasitic platyhelminth *Schistomosa mansoni* [71] and further confirmed in other animal species, including protostomes and deuterostomes (both non-chordate and chordates; see [72]), suggesting an origin that predates the protostome-deuterostome split. Using the metazoan-scale orthology analysis performed in this study, the two sequences identified in *L. stagnalis* belong to the same orthogroup (OG00004255), which is composed of 33 sequences from molluscan (Gastropoda and Bivalvia), annelid, and platyhelminth taxa, plus the hemichordate *Saccoglossus kowalevskii*. However, in this OG, no protein other than the two *L. stagnalis*' possess 2 DBDs, which may reflect conflicts in orthology search due to this very particular NR structure. Although a stable phylogeny of 2DBD-NRs is still to be established, Wu and LoVerde [72] proposed to group them into a new family (NR7). Accordingly, we found higher sequence similarity among the four DBDs of these two 2DBD-NRs than with any other *L. stagnalis* NR or with their corresponding best hits (see green clade in Fig. 1 and Supplementary Figure SF2).

To take the analysis of the unusual 2DBD-NRs further, we concatenated the DBD and LBD domains from 60 other 2DBD-NRs representative of the metazoan groups studied by Wu and Loverde [72] (see Supplementary Table ST4). A phylogenetic analysis of these domains suggests that the two *L. stagnalis* receptors are members of two different groups (A vs B; Fig. 2 and Supplementary Figure SF3), albeit with

low statistical support (see branch bootstrap values). Therefore, following the recommendation of Wu et al. 2023, *L. stagnalis* has two NR7 genes, GSLYST00003556001 (NR7A) and GSLYST00001311001 (NR7B), corresponding respectively to *C. gigas* 2DBDγ and 2DBDδ [73].

Among the incomplete NRs, four genes were found to lack either the entire LBD (1 gene) or the DBD (3 genes) and should thus be classified as members of the NR0 family (although this grouping is not monophyletic). We tentatively annotated them following their BLAST best-hit (*H. sapiens* and/or *nr*). On this basis, the single DBD-only gene was annotated as NR1A2_THRB, despite the lack of a true thyroid system in Mollusca [but see 74], and the three LBD-only genes as NR1D2_REVERBB, NR1H3_LXRA, and NR2F1_COUP-TFI, respectively. Note that NR1D2 and NR1H3 annotations were shared by at least one other gene prediction which contained both the DBD and LBD. Another specific case was GSLYST00019008001 (and its related transcript). Although we could not find an open reading frame (ORF) that would generate both the DBD and LBD, two alternative ORFs would generate sequences with either of the domains. Ambiguously, each of these sequences has a different BLAST best-hit in *H. sapiens*: NR1C1_PPARA for the DBD-only ORF, and NR1H3_LXRA for the LBD-only ORF. In addition to the observation that NR1C1 and NR1H3 are already represented by two and three other gene predictions respectively, this suggests that GSLYST00019008001 and its corresponding transcript do not constitute a functional NR.

Despite extensive searches for alternative ORFs, 5 NRs were found to have an incomplete domain sequence (lacking part of either the DBD or the LBD). These might represent non-functional gene copies, as suggested by the observed density of nucleotide repeats in the problematic areas. While this should be confirmed through specific re-sequencing, the fact that some of these and other genes with complete domains share the same functional annotation (see NR1C3_PPARG and NR2E1_TLL, Supplementary Table ST2) tends to support the hypothesis of non-functionality. One last specific situation is that of NR2E3 (PNR), which is represented by two protein predictions, neither of which have both domains in full (an incomplete DBD, and the LBD is either complete or missing).

G protein-coupled receptors

InterProScan identified a significant GPCR hit for a total of 892 gene predictions, four of which were discarded due to incompleteness or other conflicting annotations (see Supplementary Table ST5). The remaining 888 gene predictions are distributed among the five main GPCR families (GRAFS system [75]) and the ancestral cAMP [76]: class A - Rhodopsin-like (784); class B – adhesion/secretin (65); class C – Glutamate (31); class E – cAMP (3); class F – Frizzled (5). Of these, 418 were members of five orthogroups (OGs) found to be expanded within the *L. stagnalis* genome (orthology analysed at metazoan level, Z-score > 2, Table 3; see also Supplementary Information for further details), while the remainder (470) was distributed across 135 non-expanded OGs. Most gene predictions belonging to the five expanded GPCR OGs were identified as Rhodopsin-like (class A) 7 trans-membrane (7TM; IPR17452). These five expanded OGs encompassed a high proportion (89 to 100%) of molluscan proteins from the

six taxa included in the orthology analysis, and four of these OGs were predominantly exclusive to gastropods (85 to 100%; Table 3).

Table 3

Summary statistics for GPCR Orthology Groups, with a focus on molluscan representatives. Percentage values indicate for each orthogroup the proportion of proteins belonging either to Mollusca, Gastropoda, or to *L. stagnalis*. Abbreviations: OG, Orthogroup. A Z-score value (number of standard deviations from the mean) > 2.0 indicates significant expansion.

OrthoGroup	InterProScan	OG size	Mollusca	Gastropoda	*L. stagnalis*	Z-score
OG0000008	Rhodopsin_R (peptide, FMRFamide)	1,215	1085 (89%)	1035 (85%)	369 (30%)	2.537
OG0001159	Rhodopsin_R (peptide, FMRFamide)	72*	71 (99%)	71 (99%)	24 (33%)	2.336
OG0002276	Rhodopsin _R (peptide, FMRFamide)	49*	48 (98%)	48 (98%)	17 (35%)	2.452
OG0002279	*Rhodopsin_R (2 GPCRs)*	*49*	*49 (100%)*	*49 (100%)*	*22 (45%)*	*3.106*
OG0002601 Absent in Deuterostoma and Arthropoda	Rhodopsin_R (peptide, FMRFamide)	45	38	22	4	< 2
OG0010705	Rhodopsin (peptide, FMRFamide)	8	7	7	4	< 2
OG0000001	Rhodopsin_R (peptide)	3,444	911	561	139	< 2
OG0004949	Rhodopsin_R (biogenic amine)	30	30 (100%)	27 (90%)	14 (47%)	3.347
OG0000121	Rhodopsin_R (peptide and biogenic amine)	286	268	115	29	< 2
OG0000012	Adhesion_R	1,072	310	186	42	< 2

*in OG0001159, the single non-molluscan member was *B. floridae*, in OG0002276 *C. elegans*.

7TM-class A Peptide GPCRs: a predominance of FMRFamide R-like

Three of the five expanded OGs with a GPCR hit (OG000008, 0G0001159, OG0002276) consisted exclusively of Rhodopsin-like receptors from the group of (7-transmembrane-domain) 7tmA_FMRFamide_R-like receptors (peptide), representing in total 424 *L. stagnalis* proteins. However, OG0002279, consisting of 49 exclusively Gastropoda proteins, was heterogeneous with only 2 GPCR hits for *L. stagnalis* proteins. The 22 other *L. stagnalis* proteins in this OG included BLASTp annotations against bacterial proteins, casting doubt on the statistical reliability of this OG. By contrast, OG0002601, although not expanded in *L. stagnalis*, was clearly an FMRFamide receptor OG (45 proteins), apparently absent in Deuterostomia and Arthropoda. Molluscs were the most represented taxa in this group, with 38 proteins, 22 of which belong to Gastropoda, and 4 to *L. stagnalis*. OG0010705 is also a small, non-expanded, FMRFamide receptor OG (8 proteins), with 7 molluscan representatives, all gastropods from the Heterobranchia order. However, the fact that one protein from this OG belongs to the arthropod *F. candida* questions the validity of this small group as a true OG. Finally, an extra-set of 44 *L. stagnalis* gene predictions from various and heterogeneous OGs were annotated as FMRFamide_R-like, although only 18 gene predictions had a complete 7TM motif (as determined with NCBI Conserved Domains database search tool). In summary, about 450 GPCRs could be confidently designated as FMRFamide_R-like. For completion, an OG of biogenic amine GPCRs (G0004949), which was also expanded in *L. stagnalis*, contained 13 receptors to muscarinic acetylcholine and one receptor to adrenaline.

We hypothesised that FMRFamide receptor proliferation may have supported the evolution of simultaneous hermaphroditism in gastropods. To test this hypothesis, we refined the analysis of orthology by focusing on gastropods, and used available genomic resources to balance hermaphroditic species (*A. californica, B. glabrata, Bulinus truncatus, Candidula unifasciata, Elysia chlorotica, L. stagnalis, Radix auricularia, Lottia gigantea*) against gonochoristic species (separate genders, *Batillaria attramentaria, Gigantopelta aegis, Haliotis rubra, Littorina saxatilis, Pomacea canaliculata, Potamopyrgus antipodarum*) (see Supplementary Information, Orthology analyses). Note that *L. gigantea* is a protandrous sequential hermaphrodite while all other species are simultaneously hermaphroditic. This analysis generated nearly 29,500 OGs (Fig. 3; Supplementary Table ST6). Because all simultaneous hermaphrodites in our analysis are in the Euthyneura infraclass of molluscs, we also identified various OGs specific to this clade, including FMRFamide receptors, c-type lectins, cytochrome P450 cyclodipeptide synthase-associated, (see Supplementary Table ST7 for domain annotation of the main euthyneura-specific OGs).

The search for FMRFamide-R among *L. stagnalis* orthologs retrieved 483 genes distributed in 89 different OGs. FMRFamide-R groups previously found expanded in *L. stagnalis* appeared no more enriched, which is not surprising as all members of the analysis were gastropods. We excluded *Radix auricularia* from this analysis because that dataset was abnormally small (less than 18,000 protein predictions in total), and that species was recurrently absent from OGs otherwise well represented by other Hygrophila species (casting further doubt on the completeness of its predicted proteome). Excluding *R. auricularia*, the number of FMRF-R amide genes was significantly higher (ANOVA F-value [1,11] = 10.21, P = 0.009) in hermaphrodites (simultaneous and sequential: 289.86 ± 149.10 genes) than in gonochoric gastropods (89.17 ± 36.88 genes). After excluding *L. gigantea* (a sequential hermaphrodite), simultaneous

hermaphrodites were more significantly enriched in FMRFamide receptor genes (329.17 ± 117.04) than non-hermaphrodite species (ANOVA F-value [1,10] = 22.95, P < 0.001).

Peptide GPCRs: a high level of ligand diversity

OG0000001 is comprised of peptide GPCRs and is the second largest OG determined by OrthoFinder with 3,444 genes representative of all taxa included. Although not expanded in *L. stagnalis*, this OG contains 139 *L. stagnalis* gene predictions, all assigned as peptide GPCRs, and predicted to bind a great diversity of molecules (more than 40 categories; see supplementary Table ST8). OG0000121 grouped gene predictions which shared four annotations with OG0000001 (in bold in Supplementary Table ST8) or were identified as other peptide receptors (neurotensin, somatostatin, FF neuropeptide, NPY peptide, opioid).

Among the rhodopsin-like gene predictions, a structural pattern was observed in the genome, with a spatial clustering of similar genes. Whether all members of a given cluster are functional genes or pseudogenes cannot be determined based on our data, yet it is to be noted that they do not all possess a complete 7TM motif and these clusters were also often interspersed with repetitive regions. In line with its high level of occurrence, FMRFamide_R-like was either the main or exclusive annotation of 30 of the 46 largest groups (size: 4 to 26 tandem GPCR genes), and also appeared in other groups. Among these, the proportion of transmembrane domains with at least 6 helices varied from 50 to 100%. Furthermore, 10 of the 30 groups of FMRFamide_R-like genes contained genes which also exhibited an additional viral chemokine GPCR hit (PHA02638 or PHA03087). Details on cluster annotations are given in Supplementary Table ST9.

Consistent with the levels of diversity observed in metazoan GPCR classes, classes B, C, E and F were more moderately represented in *L. stagnalis*. After discarding gene predictions with transmembrane domains containing less than 6 helices (considered as incomplete), we found 37 proteins representative of class B (12 B1_hormone receptors, 19 B2_adhesion, and 6 B3_Methuselah), 17 of class C (3 GABA B, 12 metabotropic glutamate, and 2 orphan GPR158), a single class E (cAMP), and five class F receptors (four Frizzled and one Smoothened protein) (see Supplementary Table ST10). Several other predicted proteins with significant matches to GPCRs from either of these classes were also identified but discarded due to incomplete or total lack of a transmembrane domain.

Spatial expression of FMRFamide receptors

While the association between FMRFamide receptor expansion and simultaneous hermaphroditism is strong (as indicated by OG composition), it is problematic that the Euthyneura are monophyletic and are also the only simultaneous hermaphrodites used in these analyses. Therefore, to clarify whether some of these receptors are associated with the hermaphroditic condition in *L. stagnalis* (as opposed to other euthyneuric traits such as detorsion), we searched for differences in the expression of receptors in our transcriptomic datasets, and accordingly a selection was used for *in situ* expression analysis. The differential expression of some FMRFamide receptors between male (prostate) and female (albumen) reproductive organs drove our selection of 4 of these receptors (GSLYST00019397001,

GSLYST00001331001, GSLYST00016909001, GSLYST00019383001). In addition, we investigated the expression patterns of three differentially expressed “traditional” nuclear receptors (GSLYST00008490001, GSLYST00009627001 and GSLYST00020139001) and the two 2DBD-NRs (GSLYST00001311001 and GSLYST00003556001). We also included *ovipostatin* (GSLYST00000615001) as a reference because it is known to be differentially upregulated in the prostate gland relative to the albumen gland [77]. The spatial expression profiles for these genes were striking in their similarity in early veligers; in general, these genes were robustly expressed in ectodermal cells of the foot, head and mantle edge (Fig. 4). The albumen and prostate glands are fully differentiated organs with separate reproductive functions, and we were able to detect qualitatively distinct spatial expression patterns for these genes in one or both reproductive organs. As expected, o*vipostatin* generated a strong signal in sections of the prostate gland, while effectively no signal was detected in the albumen gland (Fig. 4). In addition, the expression patterns of the two 2DBD-NRs in early embryos (early and late blastulae, early and late gastrulae) indicates that these gene products play likely critical roles in early developmental processes (Fig. 5).

Discussion

Our characterisation of the genome of *L. stagnalis* expands the available range of organisms to study molluscan biology in the genomic era and provides a fundamental resource for research involving *L. stagnalis* in its role as a longstanding, invertebrate model organism. Specifically, our analysis revealed that the nuclear genome of *L. stagnalis* encodes an abundance of receptors, including DNA-binding-, 2DBD-, RTK-receptors and GPCRs. We focused our attention on the identification of receptors involved in biological processes because such information is lacking in many fields that employ *Lymnaea* as a model. For example, in ecotoxicology mechanistic evidence for toxicity pathways is critically lacking in *L. stagnalis*, even though it is used as a model for its ecological value (functional representativity within food-webs and in other ecological interactions). The resources provided in this work will be of particular interest for the study of endocrine disruptors (nuclear receptors, domain annotation and phylogeny) and more generally for xenobiotic responsive pathways involving GPCRs (e.g., [78]) or RTKs [65].

We have focused our analysis of the *L. stagnalis* genome on the identification of molecular receptors, and the expression, evolution and diversification of a selection of these. We have also characterised the mitogenome (see Supplementary Information), two unique double DNA-binding domain (2DBD) containing genes [72], and investigated the *in situ* expression profiles of several receptors during embryogenesis and in sex-specific adult organs of this simultaneous hermaphrodite. These findings complement a previous study that concentrated on ion-channels and ionotropic receptors in the central nervous system (CNS, [67]). Perhaps most notably, we observe an association between the evolution of simultaneous hermaphroditism and the diversification of GPCRs.

Identification of potential male/female receptors

As *L. stagnalis* is a simultaneous hermaphrodite, meaning each individual can execute male and female reproductive processes (see below), we searched for receptors that are potentially involved in regulating this simultaneous state. We used RNAseq data from albumen gland (AG; female) and prostate glands (PG; male) to look for sex-specific expression patterns. Indeed, some GPCRs and NRs were exclusively expressed in one or the other of these organs (Supplementary Table ST11). These potential sex-specific receptors are particularly interesting because these are potential targets of seminal fluid proteins (LyACPs, Lymnaea Accessory Gland Proteins) which have been shown to specifically influence male or female reproductive processes in the sperm recipient (e.g. [55, 54]). In the future, it will be fruitful to test the regulatory status of these receptor genes, how transient their expression is, and whether this is influenced by mating status (e.g., recently inseminated). We expect that such efforts will not only result in de-orphanising these receptors but will also present new opportunities for understanding the proximate mechanisms via which LyACPs induce their effects. This also argues for the generation of more well-annotated chromosome level gastropod genomes, and for the functional analysis of specific gene products.

Receptor expression

The broadly similar spatial expression patterns of the 2 NRs, 3 GPCRs and *ovipostatin* in early veliger larvae is intriguing (Fig. 4), and hints at potential functional commonalities for these distinct genes at this stage of development. These similarities are made more remarkable by the fact that *ovipostatin* is not a receptor, but rather a secreted accessory gland protein produced by the prostate gland that is understood to repress oviposition behaviours in *Lymnaea* [77]. These genes expand their expression from an apparently exclusive ectodermal pattern in early veligers, to the endodermally-derived adult reproductive organs we investigated (Fig. 4). The presence of mRNA transcripts for the selected receptors in one or both of these terminally differentiated organs implies that they are functionally required in these contexts, and that they therefore contribute to the maintenance of simultaneously functioning male and female reproductive states. As mentioned above, gene-specific (and likely reproduction-condition-specific) assays are required to elucidate the functional roles of these genes in these contexts. However, these results provide a foundation from which future investigations can be started.

The expression of the 2DBD-NR *retinoic acid receptor-"beta"* (as referred to in [73]) gene (GSLYST00001311001) in distinct cells of gastrula embryos (Fig. 5) corresponds with earlier reports on the effect of lithium (which is known to interact synergistically with retinoic acid [79]) on gastrulation and development in *L. stagnalis* [80]. The expression pattern of the 2DBD-NR *subfamily 1 group D* gene (GSLYST00003556001) in numerous nuclei of gastrula embryos also suggests an important role in early development, however, pending functional data (such as knock-downs or knock-outs) the specific ontogenetic roles of these genes remain unknown.

GPCR evolution/expansion

In general, G protein-coupled receptors (GPCRs) are an important class of membrane receptors that are involved in many biological processes, including external (e.g., pheromones, ions, odours, toxins, taste,

and light) and internal (e.g., neuropeptides, proteins, lipids, neurotransmitters, and nucleotides) sensory signal detection (e.g. [81]). As a superfamily, GPCRs have evolved by tandem local duplication during metazoan evolution (prior to two rounds of whole-genome duplication in vertebrates [82]). In this context it is significant that we found around 450 FMRFamide R-like (Rhodopsin-like) GPCRs in the *Lymnaea stagnalis* genome. The presence of this expanded set of receptors can be explained by the fact that they are the receptors for the many FMRFamide-like (neuro)peptides that are involved in regulating many biological processes, including the hermaphroditic condition (e.g. [83]). These peptides belong to a class that contain the C-terminal Arg-Phe-NH2 (RFamide; RFamide-related peptides or RFRPs). The evolutionary origins of these (neuro)peptides dates back to a common bilaterian ancestor [84]. The first of these peptides to be discovered was the cardio-excitatory tetrapeptide FMRFamide, identified in the neural ganglia of the bivalve mollusc, *Macrocallista nimbosa* [85]. Since then, FMRFamide has been reported to be present in many species, including the ecdysozoan protostome *Caenorhabditis elegans* [86], and various lophotrochozoan molluscs, including Cephalopods [87, 88] such as *Sepia officinalis* [89] and *Octopus vulgaris* [90, 91], and Gastropods including *Cornu aspersum, Biomphalaria glabrata* and *Biomphalaria alexandrina* [92, 93], to name but a few. In addition, many RFRPs have since been discovered and are involved, besides cardiovascular regulation, in osmoregulation, digestion, locomotion, feeding and reproduction (e.g. [94]).

The mechanism that supported the apparent expansion of these receptors in the Euthyneura remains unknown. Whole genome duplication (WGD) events have been identified within several molluscan clades (at the base of the Cephalopoda, and within the Stylommatophora, the Neogastropoda, Tonnoidea, Cypraeidae, Capulidae and Heterobranchia) but apparently not at the origin of Euthyneura [95–97]. With respect to GPCRs, due to the high level of gene clustering observed in *L. stagnalis*, it seems unlikely that a WGD event at the origin of the Euthyrneura would be responsible for the FMRFamide receptor proliferation we observe in euthyneuran species. However, this hypothesis should be tested further, using appropriate tools and chromosome-level datasets in the Mollusca and elsewhere in the animal kingdom. Interestingly Rondón et al. [98] recently reported an expansion of GPCR genes in two lineages of caenogastopods, one of which belongs to a lineage that underwent a whole genome duplication. As in the *Lymnaea* genome they also observe clustering of expanded GPCRs, which argues against a genome duplication event being a direct explanation to the observed expansion in GPCRs.

Based on our analyses of molluscan RFRPs and GPCRs, we postulate that the patterns of their abundance may in part have supported the transition from a separate-sexed state to a hermaphroditic condition. Our phylogenetic analysis supports an expansion of GPCRs in parallel with the occurrence of hermaphroditism in the molluscan species with well annotated genomes that were available at the time of writing for comparison. Supporting this observation is the fact that many molluscan reproductive processes are regulated by RFRPs and other (neuro)peptides. As in vertebrates, separate sexed molluscs can use the same substances for regulating male and female reproductive processes, because each body only contains one sex (e.g., *Octopus vulgaris* [90]; *Sepia officinalis*: [89]; *Sepiella japonica*: [99]; *Haliotis discus hannai*: [100]). However, the evolution of simultaneous hermaphroditism means that male and female

processes must be regulated concurrently within one body (*e.g., Lymnaea stagnalis, Cornu aspersum, Aplysia californica* [87]). This requires largely nonoverlapping neurobiological wiring, regulatory peptides and accompanying receptors to avoid the simultaneous expression of conflicting reproductive processes [101]. Hence, excitatory and inhibitory crosstalk between the male and female regulatory systems needs to be accompanied by an expansion of the receptors involved, so that they can be mutually exclusively used for executing one sexual role (while potentially inhibiting the other role).

It is worth illustrating this with the multitude of roles that FMRFamide and its co-expressed peptides fulfil [102–105] next to being cardio-excitatory [106]. These tetra- and heptapeptides have been shown to be important for the fine coordination of the execution of male mating behaviour and insemination [105, 107]. Importantly, FMRFamide has also been shown to inhibit the activity of the caudo-dorsal cells (CDC), which are normally responsible for the release of egg laying hormone during their synchronous bursting (CDCH [108, 109, 110]). Interestingly, the single, previously identified receptor for an RFamide is a ligand for the decapeptide TPHWRPQGRF-NH_2 (LyCEP/luqin/tachykinin [111]). This peptide is cardioexcitatory, but also inhibits CDC activity, and by extension egg laying. Thus, when the CDCs are activated and egg laying is triggered, LyCEP release is likely suppressed, which would agree with our observation of a reduction in heartrate during egg laying (Smelik et al. unpubl.). This illustrates how the different RFRPs are interlinked and can collectively stimulate one physiological function while simultaneously inhibiting conflicting functions.

Conclusion

Here we report a full genome assembly and annotation for the great pond snail *Lymnaea stagnalis*. This snail has long served as a model organism to a wide variety of scientific disciplines, and we envisage this resource will assist traditionally distinct communities to more efficiently collaborate, provide deeper insight into long-standing questions, and to stimulate novel research directions. The annotation of this genome highlights a remarkable expansion of receptors critical for a wide variety of biological processes, and hints at a causal link with the evolution of simultaneous hermaphroditism in gastropods. The identification of such receptors is also crucial for understanding the chemosensory lifestyle that many gastropods lead and will likely provide insight into the mechanisms that underlie phenomena such as pheromone release, egg-laying, predator detection, mate location, mate-partner recognition, avoidance behaviours, memory, parasite infection and co-habitation behaviours. We envisage that our ongoing efforts to identify novel receptors, in combination with parallel advances in structural *in silico* modelling of protein structure, protein receptor localisation techniques and functional studies of receptors will ultimately de-orphanise the many receptors present in the genome of *L. stagnalis and* will provide insight not only into their mechanisms of action, but into their evolutionary histories.

Materials and Methods

Snail line generation

The individual snail used for genome sequencing and assembly was derived from an inbred line previously used to map and identify the snail chirality locus [45]. This line was derived by alternating rounds of self-fertilisation and full-sib mating, over at least ten generations. The inbred snail line is freely available from the authors upon request.

Extraction of high molecular weight DNA

One whole adult snail was flash-frozen and stored at -80°C. Upon thawing, the soft tissue was removed from the shell, cleaned in a physiological water bath and ground with a mortar and pestle under liquid nitrogen. The ground material was resuspended in 15 mL SEB1X buffer in the presence of 0.2% β-mercaptoethanol and gently stirred for 15 min at 4°C. This solution was then passed through a strainer in order to remove as much mucus as possible. 0.05 vol of a SEB1X-Triton mixture was then added to the 14 mL of filtrate, incubated for 10 min at 4° C, and centrifuged for 20 min at 600g at 4° C. The supernatant was aspirated, and the resulting pellet of nuclei was washed 2 times with SEB1X buffer, then resuspended in 300 µL of S buffer in the presence of 30 µL of SDS. Buffer composition is available as Supplementary Information.

The protocol was immediately continued by embedding the nuclei in plugs. Low melting point agarose (2%) was prepared and added to the nuclei suspension, volume by volume. After gentle homogenisation with a vortex, avoiding bubbles, the mixture was poured into plugs at a rate of 100 µL per plug and left to set on ice for 15 min. The plugs were then transferred into a 50 mL tube in the presence of 10 mL lysis buffer per plug and incubated for 1h at 37°C. After removal of the lysis buffer, plugs were incubated in 10 mL of digestion buffer and proteinase-K at 40 µg/mL final concentration at 50°C overnight. The following day, the proteinase-K solution was removed and inactivated by washing with a mixture of 10 mL TE 10:1 pH 8 and 23 µL PMSF (phenylmethylsulphonyl fluoride) 0.1 M (40 µg/mL final). Two washes of 30 min at room temperature were followed by 3 washes for 10 min each at room temperature with 20 mL of TE 10:1 pH 8. The plugs were then stored in 30 mL EDTA 0.5 M pH 8 at 4°C until extraction.

High-molecular-weight DNA was extracted from the plugs by β-agarase treatment. After 5 washes in 20 mL TE 1X buffer followed by 2 washes in 20 mL MES buffer, each plug was briefly dried on paper, placed in a 1.5 mL tube and incubated for 7 min at 70°C in a dry water bath before being incubated at 42°C overnight in the presence of 16 µL β-agarase. The following day, the entire volume (~ 500 µl) was transferred onto a dialysis filter using a broad tip and left for 30 min at room temperature. DNA quality was checked using Opgen QCards (Opgen, Rockville, MD, USA), and concentration determined via Qubit Fluorometric Quantitation (Thermo Fisher Scientific, Waltham, MA, USA).

A set of cDNA libraries were constructed using various tissues and organs from individuals of the same strain used for genome sequencing (see Supplementary Table ST12). For the stomach, digestive gland, reproductive system, and foot, stranded mRNAs were generated using Illumina® TruSeq stranded mRNA Library Prep, whereas for nervous ganglia, head and mantle tissue, a kit adapted to low input RNA was used (SMARTer® Universal Low Input RNA Kit).

Transcriptome assembly

Illumina reads were quality filtered and trimmed using the procedure described in Alberti et al. [112]. In addition, for RNA-Seq data, ribosomal RNA-like reads were detected using SortMeRNA [113] and Oases 0.2.08 [114], using a k-mer size of 63 bp. Reads were mapped back to the contigs with BWA-mem and coverage was computed based on solely consistent paired-end reads. Uncovered regions were extracted and used to identify hypothetic chimeric contigs. In addition, open reading frames (ORF) and domains were searched using respectively TransDecoder (Haas, BJ. https://github.com/TransDecoder/TransDecoder) and CDsearch [115]. Uncovered regions that do not contain an ORF or a domain were tagged as mis-joins and contigs were split. Read strand information was used to correctly orient RNA-Seq contigs.

Error-correction of cDNA nanopore reads

Raw nanopore reads were corrected using Illumina RNA-Seq reads and the NaS software [116] with the following parameters: –ilmn_size 150 –mode fast –nb_proc 4 –t 2 –k 32 –untgl_seq_size 300. As the computation time is correlated with the number of Illumina reads, we decided to reduce the complexity by adding a normalisation step of the Illumina short-reads using the normalise-by-median.py script of the khmer package [117] with the following parameters: -C 20 -k 20 -N 4 -x 10e9.

DNA sequencing and genome assembly

Library preparation and whole-genome sequencing

Four Illumina PE libraries were prepared starting from 100 ng to 250 ng DNA. Four independent DNA fragmentations were performed using the E210 Covaris instrument (Covaris, Inc., USA) in order to generate fragments mostly around 400 and 500 bp. Libraries were constructed using the NEBNext DNA Sample Prep Master Mix Set (New England Biolabs, MA, USA). DNA fragments were PCR-amplified using Platinum Pfx DNA polymerase (Invitrogen) and P5 and P7 primers. Amplified library fragments were size selected on 3% agarose gel around 400 bp or on 2% agarose gel around 600bp and 700 bp. Library traces were validated on Agilent 2100 Bioanalyzer (Agilent Technologies, USA) and quantified by qPCR using the KAPA Library Quantification Kit (KapaBiosystems) on a MxPro instrument (Agilent Technologies, USA). The paired end (PE) libraries were sequenced with 250bp paired-end reads on an Illumina HiSeq 2500 instrument (Illumina, San Diego, CA, USA).

The Mate Pair (MP) libraries were prepared using the Nextera Mate Pair Sample Preparation Kit (Illumina, San Diego, CA). Briefly, genomic DNA (4 µg) was simultaneously enzymatically fragmented and tagged with a biotinylated adaptor. Tagmented fragments were size selected on a Sage Science Electrophoretic Lateral Fractionator or SageELF (Sage Science, MA, USA). This system allows production of narrowly sized MP libraries, isolating 12 different discrete size fractions from a single sample loading. We selected 8 fractions (from 3Kb to 15 Kb) to continue the MP library preparation. Size-selected fractions were circularised overnight with a ligase. Linear, non-circularized fragments were digested, and circularised DNA was fragmented to 300-1000-bp size range using Covaris E210. Biotinylated DNA was immobilised

on streptavidin beads, end-repaired, 3`-adenylated, and Illumina adapters were added. DNA fragments were PCR-amplified using Illumina adapter-specific primers and purified. Finally, libraries were quantified by qPCR and libraries profiles were evaluated using an Agilent 2100 bioanalyzer (Agilent Technologies, USA). Each library was sequenced using 100 base-length read chemistry on a paired-end flow cell on the Illumina HiSeq2000 (Illumina, USA).

Genomic DNA library coverage is summarized in Supplementary Table ST13.

Genome profile and assembly

K-mers were counted with jellyfish1 [118] to draw the genomic profile. Illumina paired-end and mate-pair reads were assembled in contigs using spades v. 3.5.02 [119] with kmer = 127. Contigs larger than 500 bp were assembled in scaffolds with sspace2.03 [120] and gaps were filled with gapfiller4 [121]. Scaffolds longer than 2 kb were kept for genome annotation.

Genome masking, repeats and transposable elements

A strategy was developed to predict and then mask repetitive DNA before gene annotation. Due to the characteristics of the obtained assembly (~ 943 Mbps; 6640 scaffolds), Transposable Elements (TEs) were *de novo* annotated with a method specifically designed to overcome computational challenges of TE discovery in large genomes and using an automated process, as implemented in the package REPET V2.5 [122, 123]. The process started with focusing on a subset of the genome sufficient to identify “easy-to-find” TEs. In brief, we eliminated stretches of *n*s > 11 from the scaffolds and kept contigs larger than 16 kbps. The resulting subset contained ~ 655 Mbps (representing 69.44% of the initial genome) and 16475 contigs, further used as effective DNA sequences to build TE consensus without N stretches.

The *de novo* detection was first self-aligned onto the genome subset and blasted against itself using the TEdenovo pipeline similarity approach implemented in REPET and based on Blaster default parameters. Matching sequences were clustered using programs specific to interspersed repeats (Grouper, Recon, Piler; see REPET package). A consensus sequence was then built for each cluster, based on the multiple alignment of all members of that cluster. All consensus sequences were then classified using the classification system of Wicker et al. [124], as implemented in PASTEClassifier tool in REPET package. Redundancy was removed and TEs were grouped by family, resulting in a filtered library of 5085 non-redundant consensus sequences representative of “easy-to-find” TEs.

TE annotation was performed using an iterative process with REPET (Blaster, Censor, RepeatMasker, see REPET package). First, the complete “easy-to-find” TE library was aligned by similarity searching on the initial genome (943 Mbps), obtaining a TE coverage of 42.6%. This led to the retrieval of 23211 TE full length copies (FLC, i.e., TE annotations aligned on more than 95% of their cognate consensus). From these, we used the corresponding consensus sequences and built a FLC TE sub-library made of 2643 sequences. Next, the FLC TE library was aligned against the initial genome (943 Mbps). Despite lower coverage (37.89% of the assembly), this alignment identified a larger number of FLCs (24190) than did

the previous one. Repeats detected *ab initio* and annotated TEs, representing 326 Mbps in total, were masked on the genome for further annotation.

The genome was first self-aligned and blasted against itself (similarity approach, using Blaster default parameters). Matches were clustered using programs specific to interspersed repeats (Grouper, Recon, Piler). A consensus sequence was built for each cluster, based on the multiple alignment of all sequences thereof. All consensus sequences were classified using the classification system of Wicker et al. [124], as implemented in PASTEClassifier. Redundancy was removed and TEs were grouped by family, resulting in a filtered library of 5085 non-redundant consensus sequences (easy-to-find TEs).

TE annotation was performed using an iterative process with TEannot pipeline (Blaster, Censor, RepeatMasker, see REPET package). First, the complete library of easy-to-find TEs was aligned by similarity searching on the initial genome (943 Mb), with a coverage of 42.6%. This led to the retrieval of 23,211 full length copies (FLC, i.e., TE annotations aligned on more than 95% of their cognate consensus). Next, the sub-library of FLC consensus sequences (n = 2,643) was aligned against the initial genome (943 Mb). Despite lower coverage (37.89%), this alignment identified a larger number of FLCs (24,190) than the previous one. Summary statistics on TE annotation are given in supplementary Tables ST14 and ST15.

Repeats detected *ab initio* and annotated TEs, representing 326 Mb in total (34.6% of the assembly), were masked on the genome for further annotation. Gene annotation was based on the search for exon/intron structure on the genome assembly and using both *ab initio* (using SNAP) and evidence driven (RNAseq) gene prediction.

Genome annotation

The general annotation workflow is summarised on Supplementary Figure SF4. Following genome masking (described above), gene prediction was performed using both *ab initio* and evidence-driven (RNA-Seq and conserved proteins) methods. Transcript contigs (obtained as described above) were first aligned against the repeat-masked genome with BLAT. For each contig, we selected a single alignment with the highest BLAT score having at least an identity percent of 90%. Additionally, any alignment containing an intron larger than 100Kb was divided into two sub-alignments. The BLAT alignments defined genomic regions where we performed spliced alignments with the corresponding RNA-Seq contig to accurately identify splice sites. These alignments were performed using est2genome [125] and those with at least 90% identity and a length ratio of aligned contig ≥ 85% were selected.

We used error-corrected cDNA nanopore reads generated as described above to calibrate the SNAP *ab initio* gene prediction tool [126]. First, ORFs were detected in corrected nanopore reads using TransDecoder (default parameters, build r20131117), and these were mapped onto the genome assembly with the same workflow used to align RNA-Seq contigs. For each genomic locus, we selected the longest nanopore read as the representative isoform, and all the representative isoforms were used to train SNAP and to

generate a *L. stagnalis* configuration file. Finally, gene structures were predicted on the masked genome assembly using SNAP and the newly created parameter file.

Gastropod protein sequences extracted from Uniprot were mapped on the masked genome using a strategy similar to the one applied to RNA-Seq contigs. Protein sequences were first masked for low complexity using segmasker [127] and aligned using BLAT. The best match and matches with a score ≥ 90% of the best match score, were retained and alignments containing introns ≥ 100kb were split to create distinct alignments. Alignments were then refined using Genewise (default parameters), which is more precise for intron/exon boundary detection [128]. Alignments were selected if more than 80% of the length of the protein was aligned to the genome. In order to catch poorly conserved genes, we also relaunched the same workflow in which the first alignment step was performed using BLASTx instead of BLAT and only the 5 best matches per genomic locus were retained.

Gene model prediction was performed using Gmove [129]. This annotation tool combines various evidence from RNAseq and conserved protein alignments, as well as *ab initio* predictions. It produces a consensus on gene structure without any prior calibration. Gmove creates an oriented graph (with exons as nodes and introns as vertex) from which all possible paths are extracted, representing potential gene models. Coding frames consistent with the available protein evidence were searched and in cases where no protein alignment exists, the longest coding frame is selected. Single-exon genes and genes located in the intron of another gene were rejected if their coding sequence was shorter than 300 nt. Then, genes that overlap TEs on 50% of their lengths were also removed. Among the remaining genes, we retained all genes with a coding sequence greater than 300 nt as well as shorter genes with at least one match to UniProt proteins (e-value < 10−10).

We analysed the completeness of the predicted gene catalogue using BUSCO software (v5.2.2). The odb10 eukaryotic BUSCO dataset contains single-copy orthologous genes found in > 90% of known eukaryotic species' genomes [130]. Annotation of the predicted genes were done via BLAST (E-value threshold of 1E- 5) search against Nr, SwissProt and KOG databases and search against InterPro using RunIprScan (http://michaelrthon.com/runiprscan/), a command line utility for interfacing with the InterProScan server at the European Bioinformatics Institute. Protein domain annotation was completed with the Conserved Domain Database (NCBI CD search tool).

Orthology and gene family analysis

An orthology analysis of the predicted proteome of *L. stagnalis* was performed using OrthoFinder at two phylogenetic levels. First, as a broad level of investigation, a set of 19 specific proteomes was chosen from Genbank to represent a large diversity of metazoan lineages: Placozoa (*Trichoplax adhaerens*), Porifera (*Amphimedon queenslandica*), Platyhelminthes (*Schistosoma mansoni*), Mollusca (*Biomphalaria glabrata, Aplysia californica, Lottia gigantea, Crassostrea gigas, Octopus bimaculoides*), Annelida (*Helobdella robusta*), Nematoda (*Caenorhabditis elegans*), Arthropoda (*Daphnia pulex, Hyalella azteca, Folsomia candida, Drosophila melanogaster*), Echinodermata (*Apostichopus japonicus*), Hemichordata (*Saccoglossus kowalevskii*), and Chordata (*Branchiostoma floridae, Ciona intestinalis,*

Homo sapiens). Orthogroups were identified at this scale of diversity. Gene family expansion was statistically assessed in *L. stagnalis* using a Z-score based test [131, 132].

Next, a refined analysis was performed at the level of Gastropoda, with a specific focus on GPCRs, to test for the possible implication of FMRFamide receptor expansion (as inferred from the first analysis) in the evolution of simultaneous hermaphroditism in this class of molluscs. The analysis was based on 14 gastropod species, including 6 species with separate sexes (*Haliotis rubra, Gigantopelta aegis, Batillaria attramentaria, Pomacea canaliculata, Potamopyrgus antipodarum, and Littorina saxatilis*), one sequential hermaphrodite (*Lottia gigantea*), and 6 simultaneous hermaphrodites (*Aplysia californica, Candidula unifasciata, Elysia chlorotica, Bulinus truncatus, Biomphalaria glabrata, Radix auricularia*, and *Lymnaea stagnalis*). From this analysis, the gene contribution of each species to all 89 orthogroups with *L. stagnalis* members annotated as “FMRFamide_receptor-like" was summed up and used to compare the mean number of genes per species across reproductive modes (simultaneous hermaphroditism vs other; hermaphroditism vs separate sexes) using an analysis of variance.

(see supplementary information, orthology analyses).

Mitochondrial genome

Sequence similarity searches of the genome assembly identified a scaffold containing the complete mitochondrial genome of *L. stagnalis*. Constituent NaS Nanopore reads were checked to evaluate contiguity of the scaffold. The initial automated annotation of the mitogenome (MITOS; http://mitos.bioinf.uni-leipzig.de/index.py; [133] was manually checked using Snapgene software (www.snapgene.com) and applying criteria for annotation of molluscan mitogenomes as recommended by Ghiselli et al. [134]. The gene order of the mitogenome was compared among *L. stagnalis* and representative hygrophylid gastropods available in GenBank.

Whole mount in-situ hybridisation

A total of 10 genes were selected for whole mount *in situ* hybridisation (WMISH) against male (prostate gland) and female (albumen) reproductive organs. We also surveyed late developmental stages for all genes, and also gastrulation stages for the two 2DBD-NRs. These included two DNA-binding-domain containing nuclear receptors (GSLYST00001311001 and GSLYST00003556001), 4 GPCR genes (GSLYST00019397001, GSLYST00001331001, GSLYST00016909001, GSLYST00019383001) and 3 nuclear receptors (GSLYST00008490001, GSLYST00009627001 and GSLYST00020139001), and as a positive control the seminal fluid protein ovipostatin (GSLYST00000615001; [77]. Fragments of these genes were amplified from cDNA using previously described standard procedures [135] and the primers described in Supplementary Table ST16. PCR fragments were cloned into a TA vector (pGEM-T Easy) and verified via Sanger sequencing. DIG labelled anti-sense riboprobes were synthesised as previously described [135]. A paraffin Tissue Microarray was 3D printed and constructed using STL files as described

in Pazaitis and Kaiser [136]. Snails for these experiments were maintained as previously described [40]. The albumen and prostate glands of mature *L. stagnalis* were dissected out as previously described [137] and fixed for 2 hours in 0.1% glutaraldehyde in PBS with 0.1% Triton at room temperature. These tissues were then washed twice with PBS + 0.1% Tween-20, and then dehydrated through a series of ethanol with an overnight incubation in 100% ethanol. The following day the tissue was exchanged into 100% xylene and incubated for a further 24 hours before being perfused with molten paraffin overnight. Tissue samples were placed into the recipient paraffin block and allowed to set for several hours to overnight (the pairing of albumen and prostate glands from individual snails was always maintained). Microtome sections (10 µm) were collected onto coated microscope slides and allowed to adhere at 37°C overnight. Following de-paraffinisation in xylene and a rehydration series, *in situ* hybridisation was performed as previously described [135].

Snails and experimental conditions for mating observations and reproductive organ RNAseq

Ten adult snails of three months old and with a shell length of 30 mm were collected from an age-synchronised mass rearing culture at Vrije Universiteit (VU) in Amsterdam. These snails were isolated for 8 days by placing each individual in a transparent, plastic container with a volume of 625 mL filled up to 460 mL. These containers were perforated with slits on opposing sides to let water flow through and were placed in a laminar flow tank with running, low-copper water. The water temperature was 20 ± 1°C and the light conditions were 12h:12h, as per standard culturing conditions at VU Amsterdam. These snails were each fed one disk of broad-leaf lettuce, measuring approximately 19 cm^2, daily. On day seven the containers were replaced with clean ones to induce egg laying, so that this behaviour would not interfere during the mating observations on the next day. On day eight, snails were randomly placed in pairs in the same type of containers, but without slits, that were filled up to 460 mL with water of the same temperature. These containers were places on a lab bench for mating observations [6h, e.g. 138]. Once unilateral copulation was completed, the sperm donor (male-acting individual) and the sperm recipient (female-acting) were euthanised by injecting 2–3 mL of 50 mM $MgCl_2$ through the foot (needle gauge: 0.3 mm x 13 mm). The (male) prostate gland and the (female) albumen gland of donors and recipients were dissected out and snap frozen for RNAseq and differential gene expression analyses. Reads (see earlier RNAseq description) were quality controlled using FastQC and MultiQC. Genome-guided alignments of this data was performed using HISAT2, followed by a quality control of the alignment using QualiMap. Reads that mapped to annotated genes were counted using HTseq, after which specific receptors (NR and GPCR) were selected for further investigation.

Declarations

Acknowledgements

JMK and YN thank J. Mariën, C. Popelier and O. Bellaoui for technical assistance. MAC thanks M. Collinet and the INRAE-U3E staff for technical support.

Funding Information

This work was supported by the Genoscope, the Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA) and France Génomique (ANR-10-INBS-09–08). JMK and YN were funded by Open Competition Grant OCENW.KLEIN.062. DJJ was funded by Deutsche Forschungsgemeinschaft (DFG project #528314512).

Author contributions

Conceptualisation/study design: JMK, DJJ, MAC, AD, CMA, ZPF

Resources: JMK, DJJ, MAC, AD, AM, BN, VJ, JP, KL, CDS, CK, JMA, PW

Data collection/curation: JMK, DJJ, MAC, AM, BN, VJ, CDS, CK, JMA

Bioinformatic investigation: JMK, DJJ, MAC, YN, AM, BN, VJ, CMA, CDS, CK, JMA

Data analysis: JMK, DJJ, MAC, YN, AM, BN, VJ, CDS, CK, JMA

Experimental investigation: DJJ, NC, JMK

Visualisation: JMK, DJJ, MAC, CMA

Writing (original draft): JMK, DJJ, MAC

Writing (review and editing): JMK, DJJ, MAC, AD, CMA, YN, AM, BN, VJ, JP, KL, CDS, CK, JMA, PW

Funding acquisition: JMK, DJJ, MAC, AD, ZPF, PW

Data availability

The inbred line of snails is available from the authors. The Illumina and nanopore sequencing data, as well as the genome assembly and gene predictions, are available in the European Nucleotide Archive (ENA) under the project PRJEB67819. Moreover, all data are freely available from the Genoscope website: www.genoscope.cns.fr/lymnaea.

References

1. Adema, C. M. *et al.* Whole genome analysis of a schistosomiasis-transmitting freshwater snail. *Nature communications* **8**, 15451 (2017).
2. Schlosser, G. A Short History of Nearly Every Sense-The Evolutionary History of Vertebrate Sensory Cell Types. *Integrative And Comparative Biology* **58**, 301–316 (2018).

3. Robertson, H. M. Molecular evolution of the major arthropod chemoreceptor gene families. *Annual review of entomology* (2019).
4. Rigon, F. *et al.* Evolutionary diversification of secondary mechanoreceptor cells in tunicata. *BMC Evol Biol* **4**, 12 (2013).
5. Santini, S. *et al.* The compact genome of the sponge Oopsacas minuta (Hexactinellida) is lacking key metazoan core genes. *BMC Biology* **21**, 139 (2023).
6. Moggioli, G. *et al.* Distinct genomic routes underlie transitions to specialised symbiotic lifestyles in deep-sea annelid worms. *Nature* (2023).
7. Chapman, J. A. *et al.* The dynamic genome of *Hydra*. *Nature* **464**, 592 (2010).
8. Kawasawa, Y., McKenzie, L. M., Hill, D. P., Bono, H., Yanagisawa, M. G protein-coupled receptor genes in the FANTOM2 database. *Genome Research* **13**, 1466 (2003).
9. Özçelik, R., van Tilborg, D., Jiménez-Luna, J., Grisoni, F. Structure-Based Drug Discovery with Deep Learning. *Chembiochem* **24**, e202200776 (2023).
10. Hauser, A. S., Gloriam, D. E., Bräuner-Osborne, H., Foster, S. R. Novel approaches leading towards peptide GPCR de-orphanisation. *British Journal of Pharmacology* **177**, 961 (2020).
11. Civelli, O. *et al.* G protein-coupled receptor deorphanizations. *Annu Rev Pharmacol Toxicol* **53**, 127 (2013).
12. Beets, I. *et al.* System-wide mapping of neuropeptide-GPCR interactions in *C. elegans*. *Cell Reports* **42**, 113058 (2023).
13. Fodor, I., AA, H. A., Benjamin, P. R., Koene, J. M., Zsolt, P. The unlimited potential of the great pond snail, *Lymnaea stagnalis*. *eLife* **9**, e56962 (2020).
14. Ducrot, V. *et al.* Development and validation of an OECD reproductive toxicity test guideline with the pond snail *Lymnaea stagnalis* (Mollusca, Gastropoda). *Regulatory Toxicology and Pharmacology* **70**, 605 (2014).
15. Hussein, A. A. A. *et al.* Effect of photoperiod and light intensity on learning ability and memory formation of the pond snail *Lymnaea stagnalis*. *Invertebrate Neuroscience* **20**, (2020).
16. Korneev, S. A. *et al.* A CREB2-targeting microRNA is required for long-term memory after single-trial learning. *Scientific Reports* **8**, (2018).
17. Rivi, V. *et al.* What can we teach *Lymnaea* and what can *Lymnaea* teach us. *Biol Rev Camb Philos Soc* **96**, 1590 (2021).
18. Amorim, J. *et al.* Lymnaea stagnalis as a freshwater model invertebrate for ecotoxicological studies. *Science of The Total Environment* **669**, 11 (2019).
19. Brix, K. V., Esbaugh, A. J., Munley, K. M., Grosell, M. Investigations into the mechanism of lead toxicity to the freshwater pulmonate snail, Lymnaea stagnalis. *Aquat Toxicol* **106-107**, 147 (2012).
20. Bouétard, A., Besnard, A.-L., Vassaux, D., Lagadic, L., Coutellec, M.-A. Impact of the redox-cycling herbicide diquat on transcript expression and antioxidant enzymatic activities of the freshwater snail Lymnaea stagnalis. *Aquatic Toxicology* **126**, 256 (2013).

21. Côte, J. *et al.* Genetic variation of *Lymnaea stagnalis* tolerance to copper: A test of selection hypotheses and its relevance for ecological risk assessment. *Environ Pollut* **205**, 209 (2015).
22. Karakaş, S. B., Otludil, B. Accumulation and histopathological effects of cadmium on the great pond snail *Lymnaea stagnalis* Linnaeus, 1758 (Gastropoda: Pulmonata). *Environmental Toxicology and Pharmacology* **78**, 103403 (2020).
23. Seppälä, O., Çetin, C., Cereghetti, T., Feulner, P. G. D., Adema, C. M. Examining adaptive evolution of immune activity: opportunities provided by gastropods in the age of 'omics'. *Philosophical Transactions of the Royal Society B: Biological Sciences* **376**, 20200158 (2021).
24. Seppälä, O. *et al.* Transcriptome profiling of *Lymnaea stagnalis* (Gastropoda) for ecoimmunological research. *BMC Genomics* **22**, 144 (2021).
25. Feng, Z.-P. *et al.* Transcriptome analysis of the central nervous system of the mollusc *Lymnaea stagnalis*. *BMC Genomics* **10**, 451 (2009).
26. Rosato, M., Hoelscher, B., Lin, Z., Agwu, C., Xu, F. Transcriptome analysis provides genome annotation and expression profiles in the central nervous system of *Lymnaea stagnalis* at different ages. *BMC Genomics* **22**, 637 (2021).
27. van Nierop, P. *et al.* Identification and Functional Expression of a Family of Nicotinic Acetylcholine Receptor Subunits in the Central Nervous System of the Mollusc *Lymnaea stagnalis*. *Journal of Biological Chemistry* **281**, 1680 (2006).
28. Wooller, S. *et al.* A combined bioinformatics and LC-MS-based approach for the development and benchmarking of a comprehensive database of *Lymnaea* CNS proteins. *J Exp Biol* **225**, jeb243753 (2022).
29. Ivashkin, E. *et al.* Serotonin Mediates Maternal Effects and Directs Developmental and Behavioral Changes in the Progeny of Snails. *Cell Rep* **12**, 1144 (2015).
30. Ivashkin, E. *et al.* Transglutaminase Activity Determines Nuclear Localization of Serotonin Immunoreactivity in the Early Embryos of Invertebrates and Vertebrates. *ACS Chemical Neuroscience* **10**, 3888 (2019).
31. Morrill, J. B., Blair, C. A., Larsen, W. J. Regulative development in the pulmonate gastropod, *Lymnaea palustris*, as determined by blastomere deletion experiments. *Journal of Experimental Zoology* **183**, 47 (1973).
32. Raven, C. P. The Development of the Egg of *Limnaea Stagnalis* L. From Oviposition Till First Cleavage. *Archives Néerlandaises de Zoologie* **7**, 91 (1946).
33. Raven, C. P. Morphogenesis in *Limnaea stagnalis* and its disturbance by lithium. *Journal of Experimental Zoology* **121**, 1 (1952).
34. Raven, C. P., Van der Wal, U. P. Analysis of the formation of the animal pole plasm in the eggs of Limnaea stagnalis. *J Embryol Exp Morphol* **12**, 123 (1964).
35. van den Biggelaar, J. A. Timing of the phases of the cell cycle during the period of asynchronous division up to the 49-cell stage in Lymnaea. *J Embryol Exp Morphol* **26**, 367 (1971).

36. Lukowiak, K., Sunada, H., Teskey, M., Lukowiak, K., Dalesman, S. Environmentally relevant stressors alter memory formation in the pond snail *Lymnaea*. *J Exp Biol* **217**, 76 (2014).
37. Martens, K. R. *et al.* Stressful stimuli modulate memory formation in *Lymnaea stagnalis*. *Neurobiology of Learning and Memory* **87**, 391 (2007).
38. Swinton, E., Swinton, C., Lukowiak, K. Shell damage leads to enhanced memory formation in *Lymnaea*. *Journal of Experimental Biology* (2019).
39. Cerveau, N., Jackson, D. J. A survey of miRNAs involved in biomineralization and shell repair in the freshwater gastropod *Lymnaea stagnalis*. *Discover Materials* **1**, (2021).
40. Herlitze, I., Marie, B., Marin, F., Jackson, D. J. Molecular modularity and asymmetry of the molluscan mantle revealed by a gene expression atlas. *GigaScience* (2018).
41. Hohagen, J., Jackson, D. J. An ancient process in a modern mollusc: early development of the shell in *Lymnaea stagnalis*. *BMC Dev Biol* **13**, 27 (2013).
42. Abe, M., Kuroda, R. The development of CRISPR for a mollusc establishes the formin *Lsdia1* as the long-sought gene for snail dextral/sinistral coiling. *Development* **146**, (2019).
43. Kuroda, R. *et al.* Diaphanous gene mutation affects spiral cleavage and chirality in snails. *Scientific Reports* **6**, (2016).
44. Shibazaki, Y., Shimizu, M., Kuroda, R. Body handedness is directed by genetically determined cytoskeletal dynamics in the early embryo. *Curr Biol* **14**, 1462 (2004).
45. Davison, A. *et al.* Formin Is Associated with Left-Right Asymmetry in the Pond Snail and the Frog. *Curr Biol* **26**, 654 (2016).
46. Ledder, M., Nakadera, Y., Staikou, A., Koene, J. M. Dominant gingers - Discovery and inheritance of a new shell polymorphism in the great pond snail Lymnaea stagnalis. *Ecol Evol* **13**, e10678 (2023).
47. Koene, J. M., Ter Maat, A. Coolidge effect in pond snails: male motivation in a simultaneous hermaphrodite. *BMC Evolutionary Biology* **7**, 212 (2007).
48. Boer, P., Jansen, R., Koene, J., Maat, A. Nervous control of male sexual drive in the hermaphroditic snail *Lymnaea stagnalis*. *J Exp Biol* **200**, 941 (1997).
49. Moussaoui, R., Verdel, K., Benbellil-Tafoughalt, S., Koene, J. M. Female behaviour prior to additional sperm receipt in the hermaphroditic pond snail *Lymnaea stagnalis*. *Invertebrate Reproduction & Development* **62**, 82 (2018).
50. Daupagne, L., Koene, J. M. Disentangling female postmating responses induced by semen transfer components in a simultaneous hermaphrodite. *Animal Behaviour* **166**, 147 (2020).
51. Álvarez, B., Koene, J. M. Cognition and Its Shaping Effect on Sexual Conflict: Integrating Biology and Psychology. *Front Integr Neurosci* **16**, 826304 (2022).
52. Charnov, E. L. Simultaneous hermaphroditism and sexual selection. *Proc Natl Acad Sci U S A* **76**, 2480 (1979).
53. Anthes, N. *et al.* Bateman Gradients in Hermaphrodites: An Extended Approach to Quantify Sexual Selection. *The American Naturalist* **176**, 249 (2010).

54. Nakadera, Y. *et al.* Receipt of Seminal Fluid Proteins Causes Reduction of Male Investment in a Simultaneous Hermaphrodite. *Current Biology* **24**, 859 (2014).
55. Koene, J. M. *et al.* Male accessory gland protein reduces egg laying in a simultaneous hermaphrodite. *PLoS One* **5**, e10117 (2010).
56. Álvarez, B., Koene, J. M., Hollis, K. L., Loy, I. Learning to anticipate mate presence shapes individual sex roles in the hermaphroditic pond snail, *Lymnaea stagnalis. Anim Cogn* **25**, 1417 (2022).
57. Puurtinen, M., Emily Knott, K., Suonpää, S., Nissinen, K., Kaitala, V. Predominance of outcrossing in *Lymnaea stagnalis* despite low apparent fitness costs of self-fertilization. *J Evol Biol* **20**, 901 (2007).
58. Escobar, J. S. *et al.* Patterns of mating-system evolution in hermaphroditic animals: correlations among selfing rate, inbreeding depression, and the timing of reproduction. *Evolution* **65**, 1233 (2011).
59. Coutellec, M.-A., Caquet, T. Heterosis and inbreeding depression in bottlenecked populations: a test in the hermaphroditic freshwater snail Lymnaea stagnalis. Genetic load in a freshwater snail. *Journal of Evolutionary Biology* **24**, 2248 (2011).
60. http://dx.doi.org/10.1787/20777876
61. Fodor, I. *et al.* Studies on a widely-recognized snail model species (*Lymnaea stagnalis*) provide further evidence that vertebrate steroids do not have a hormonal role in the reproduction of mollusks. *Frontiers in Endocrinology* **13**, 1 (2022).
62. Bertrand, S. *et al.* Evolutionary Genomics of Nuclear Receptors: From Twenty-Five Ancestral Genes to Derived Endocrine Systems. *Molecular Biology and Evolution* **21**, 1923 (2004).
63. de Mendoza, A., Sebé-Pedrós, A., Ruiz-Trillo, I. The evolution of the GPCR signaling system in eukaryotes: modularity, conservation, and the transition to metazoan multicellularity. *Genome Biol Evol* **6**, 606 (2014).
64. Morel, M., Vanderstraete, M., Hahnel, S., Grevelding, C. G., Dissous, C. Receptor tyrosine kinases and schistosome reproduction: new targets for chemotherapy. *Frontiers in Genetics* **5**, (2014).
65. Lemmon, M. A., Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases. *Cell* **141**, 1117 (2010).
66. El Zein, N., D'Hondt, S., Sariban, E. Crosstalks between the receptors tyrosine kinase EGFR and TrkA and the GPCR, FPR, in human monocytes are essential for receptors-mediated cell activation. *Cellular Signalling* **22**, 1437 (2010).
67. Dong, N. *et al.* Ion channel profiling of the *Lymnaea stagnalis* ganglia via transcriptome analysis. *BMC Genomics* **22**, 18 (2021).
68. Frooninckx, L. *et al.* Neuropeptide GPCRs in *C. elegans*. *Frontiers in Endocrinology* **3**, (2012).
69. Vinogradov, A. E. Variation in ligand-accessible genome size and its ecomorphological correlates in a pond snail. *Hereditas* **128**, 59 (1998).
70. Elliott, T. A., Gregory, T. R. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **370**, 20140331 (2015).

71. Wu, W., Niles, E. G., El-Sayed, N., Berriman, M., LoVerde, P. T. *Schistosoma mansoni* (Platyhelminthes, Trematoda) nuclear receptors: sixteen new members and a novel subfamily. *Gene* **366**, 303 (2006).
72. Wu, W., LoVerde, P. T. Updated knowledge and a proposed nomenclature for nuclear receptors with two DNA binding domains (2DBD-NRs). *PLoS One* **18**, e0286107 (2023).
73. Vogeler, S., Galloway, T. S., Lyons, B. P., Bean, T. P. The nuclear receptor gene family in the Pacific oyster, *Crassostrea gigas*, contains a novel subfamily group. *BMC Genomics* **15**, 369 (2014).
74. Morthorst, J. E., Holbech, H., De Crozé, N., Matthiessen, P., LeBlanc, G. A. Thyroid-like hormone signaling in invertebrates and its potential role in initial screening of thyroid hormone system disrupting chemicals. *Integr Environ Assess Manag* **19**, 63 (2023).
75. Schiöth, H. B., Fredriksson, R. The GRAFS classification system of G-protein coupled receptors in comparative perspective. *General and comparative endocrinology* **142**, 94 (2005).
76. Krishnan, A., Almén, M. S., Fredriksson, R., Schiöth, H. B. The origin of GPCRs: identification of mammalian like Rhodopsin, Adhesion, Glutamate and Frizzled GPCRs in fungi. *PLoS One* **7**, e29817 (2012).
77. Swart, E. M. *et al.* Temporal expression profile of an accessory-gland protein that is transferred via the seminal fluid of the simultaneous hermaphrodite *Lymnaea stagnalis*. *Journal of Molluscan Studies* **85**, 177 (2019).
78. Kim, D.-H., Park, J. C., Lee, J.-S. G protein-coupled receptors (GPCRs) in rotifers and cladocerans: Potential applications in ecotoxicology, ecophysiology, comparative endocrinology, and pharmacology. *Comp Biochem Physiol C Toxicol Pharmacol.* **256**, 109297 (2022).
79. Holtz, K. M., Rice, A. M., Sartorelli, A. C. Lithium chloride inactivates the 20S proteasome from WEHI-3B D+ leukemia cells. *Biochemical and Biophysical Research Communications* **303**, 1058 (2003).
80. Raven, C. P. Lithium as a tool in the analysis of morphogenesis in *Limnaea stagnalis*. *Experientia* **8**, 252 (1952).
81. Heo, J. *et al.* Prediction of the 3D structure of FMRF-amide neuropeptides bound to the mouse MrgC11 GPCR and experimental validation. *Chembiochem* **8**, 1527 (2007).
82. Yun, S. *et al.* Prevertebrate Local Gene Duplication Facilitated Expansion of the Neuropeptide GPCR Superfamily. *Mol Biol Evol* **32**, 2803 (2015).
83. Kriegsfeld, L. J. Driving reproduction: RFamide peptides behind the wheel. *Horm Behav* **50**, 655 (2006).
84. Elphick, M. R., Mirabeau, O. The Evolution and Variety of RFamide-Type Neuropeptides: Insights from Deuterostomian Invertebrates. *Front Endocrinol (Lausanne)* **5**, 93 (2014).
85. Price, D. A., Greenberg, M. J. Structure of a Molluscan Cardioexcitatory Neuropeptide. *Science* **197**, 670 (1977).
86. Ubuka, T., Tsutsui, K. Comparative and Evolutionary Aspects of Gonadotropin-Inhibitory Hormone and FMRFamide-Like Peptide Systems. *Frontiers in Neuroscience* **12**, (2018).

87. Koene, J. M., Jansen, R. F., Ter Maat, A., Chase, R. A conserved location for the central nervous system control of mating behaviour in gastropod molluscs: evidence from a terrestrial snail. *J Exp Biol* **203**, 1071 (2000).
88. Rőszer, T., Kiss-Tóth, É. D. FMRF-amide is a glucose-lowering hormone in the snail *Helix aspersa*. *Cell and Tissue Research* **358**, 371 (2014).
89. Henry, J., Zatylny, C., Boucaud-Camou, E. Peptidergic control of egg-laying in the cephalopod *Sepia officinalis*: involvement of FMRFamide and FMRFamide-related peptides. *Peptides* **20**, 1061 (1999).
90. Di Cristo, C., Paolucci, M., Iglesias, J., Sanchez, J., Di Cosmo, A. Presence of two neuropeptides in the fusiform ganglion and reproductive ducts of *Octopus vulgaris*: FMRFamide and gonadotropin-releasing hormone (GnRH). *Journal of Experimental Zoology* **292**, 267 (2002).
91. Di Cristo, C. Nervous control of reproduction in *Octopus vulgaris*: a new model. *Invertebrate Neuroscience* **13**, 27 (2013).
92. Acker, M. J. *et al.* An immunohistochemical analysis of peptidergic neurons apparently associated with reproduction and growth in *Biomphalaria alexandrina*. *General and Comparative Endocrinology* **280**, 1 (2019).
93. Rolón-Martínez, S. *et al.* FMRF-NH_2 -related neuropeptides in *Biomphalaria* spp., intermediate hosts for schistosomiasis: Precursor organization and immunohistochemical localization. *J Comp Neurol* **529**, 3336 (2021).
94. Zatylny-Gaudin, C. *et al.* Characterization of a novel LFRFamide neuropeptide in the cephalopod *Sepia officinalis*. *Peptides* **31**, 207 (2010).
95. Hallinan, N. M., Lindberg, D. R. Comparative analysis of chromosome counts infers three paleopolyploidies in the mollusca. *Genome Biol Evol* **3**, 1150 (2011).
96. Farhat, S., Modica, M. V., Puillandre, N. Whole Genome Duplication and Gene Evolution in the Hyperdiverse Venomous Gastropods. *Mol Biol Evol* **40**, msad171 (2023).
97. Liu, C. *et al.* Giant African snail genomes provide insights into molluscan whole-genome duplication and aquatic-terrestrial transition. *Molecular Ecology Resources* **21**, 478 (2021).
98. Rondón, J. J. *et al.* Comparative genomic analysis of chemosensory-related gene families in gastropods. *Mol Phylogenet Evol* **192**, 107986 (2023).
99. Li, Y. *et al.* Identification, Characterization, and Expression Analysis of a FMRFamide-Like Peptide Gene in the Common Chinese Cuttlefish (*Sepiella japonica*). *Molecules* **23**, 742 (2018).
100. Sukhan, Z. P., Cho, Y., Hossen, S., Lee, W. K., Kho, K. H. Identification and Characterization of Hdh-FMRF2 Gene in Pacific Abalone and Its Possible Role in Reproduction and Larva Development. *Biomolecules* **13**, 109 (2023).
101. C. Di Cristo, J. M. Koene, in *The Oxford Handbook of Invertebrate Neurobiology. Mechanisms and Evolution*, Eds. (Oxford University Press, 2017), pp. 615-662.
102. Schot, L. P. C., Boer, H. H. Immunocytochemical demonstration of peptidergic cells in the pond snail *Lymnaea stagnalis* with an antiserum to the molluscan cardioactive tetrapeptide FMRF-amide. *Cell*

and Tissue Research **225**, (1982).

103. Bright, K. *et al.* Mutually exclusive expression of alternatively spliced FMRFamide transcripts in identified neuronal systems of the snail Lymnaea. *J Neurosci* **13**, 2719 (1993).
104. Santama, N. *et al.* Processing of the FMRFamide precursor protein in the snail *Lymnaea stagnalis*: characterization and neuronal localization of a novel peptide, 'SEEPLY'. *Eur J Neurosci* **5**, 1003 (1993).
105. van Golen, F. A., Li, K. W., de Lange, R. P., Jespersen, S., Geraerts, W. P. Mutually exclusive neuronal expression of peptides encoded by the FMRFa gene underlies a differential control of copulation in *Lymnaea*. *J Biol Chem* **270**, 28487 (1995).
106. Buckett, K. J., Peters, M., Dockray, G. J., Van Minnen, J., Benjamin, P. R. Regulation of heartbeat in Lymnaea by motoneurons containing FMRFamide-like peptides. *J Neurophysiol* **63**, 1426 (1990).
107. El Filali, Z. *et al.* Single-cell analysis of peptide expression and electrophysiology of right parietal neurons involved in male copulation behavior of a simultaneous hermaphrodite. *Invertebrate Neuroscience* **15**, (2015).
108. Brussaard, A. B., Kits, K. S., Ter Maat, A., Van Minnen, J., Moed, P. J. Dual inhibitory action of FMRFamide on neurosecretory cells controlling egg laying behavior in the pond snail. *Brain Res* **447**, 35 (1988).
109. Brussaard, A. B., Schluter, N. C. M., Ebberink, R. H. M., Kits, K. S., Maat, A. T. Discharge induction in molluscan peptidergic cells requires a specific set of autoexcitatory neuropeptides. *Neuroscience* **39**, 479 (1990).
110. Croll, R. P., J. Van Minnen, K. S. Kits, A. B. Smit, in *Molluscan Neurobiology*, K. S. Kits, H. H. Boer, J. Joosse, Eds. (North-Holland, Amsterdam, 1991), pp. 248-254.
111. Tensen, C. P. *et al.* The lymnaea cardioexcitatory peptide (LyCEP) receptor: a G-protein-coupled receptor for a novel member of the RFamide neuropeptide family. *J Neurosci* **18**, 9812 (1998).
112. Alberti, A. *et al.* Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. *Sci Data* **4**, 170093 (2017).
113. Kopylova, E., Noé, L., Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211 (2012).
114. Schulz, M. H., Zerbino, D. R., Vingron, M., Birney, E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086 (2012).
115. Marchler-Bauer, A., Bryant, S. H. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* **32**, W327 (2004).
116. Madoui, M. A. *et al.* Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC Genomics* **16**, 327 (2015).
117. MR, Crusoe *et al.* The khmer software package: enabling efficient nucleotide sequence analysis [version 1; peer review: 2 approved, 1 approved with reservations]. *F1000Research* **4**, (2015).

118. Marçais, G., Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764 (2011).
119. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455 (2012).
120. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578 (2011).
121. Boetzer, M., Pirovano, W. Toward almost closed genomes with GapFiller. *Genome Biol* **13**, R56 (2012).
122. Quesneville, H., Nouaud, D., Anxolabéhère, D. Detection of new transposable element families in *Drosophila melanogaster* and *Anopheles gambiae* genomes. *J Mol Evol* **57 Suppl 1**, S50 (2003).
123. Jamilloux, V., Daron, J., Choulet, F., Quesneville, H. De Novo Annotation of Transposable Elements: Tackling the Fat Genome Issue. *Proceedings of the IEEE* 1 (2016).
124. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8**, 973 (2007).
125. Mott, R. EST_GENOME: a program to align spliced DNA sequences to unspliced genomic DNA. *Bioinformatics* **13**, 477 (1997).
126. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
127. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, (2009).
128. Birney, E., Clamp, M., Durbin, R. GeneWise and Genomewise. *Genome Research* **14**, 988 (2004).
129. Dubarry, M. *et al.* Gmove a tool for eukaryotic gene predictions using various evidences. *Next Generation Sequencing Conference (NGS) 2016* (2016).
130. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210 (2015).
131. Cao, Z. *et al.* The genome of *Mesobuthus martensii* reveals a unique adaptation model of arthropods. *Nat Commun* **4**, 2602 (2013).
132. Faddeeva-Vakhrusheva, A. *et al.* Coping with living in the soil: the genome of the parthenogenetic springtail *Folsomia candida*. *BMC Genomics* **18**, 493 (2017).
133. Donath, A. *et al.* Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic Acids Res* **47**, 10543 (2019).
134. Ghiselli, F. *et al.* Molluscan mitochondrial genomes break the rules. *Philos Trans R Soc Lond B Biol Sci* **376**, 20200159 (2021).
135. Jackson, D. J. Mantle modularity underlies the plasticity of the molluscan shell: supporting data from *Cepaea nemoralis*. *Frontiers in Genetics* **12**, (2021).
136. Pazaitis, N., Kaiser, A. TMA-Mate: An open-source modular toolkit for constructing tissue microarrays of arbitrary layouts. *HardwareX* **14**, e00419 (2023).
137. van Iersel, S., Swart, E. M., Nakadera, Y., van Straalen, N. M., Koene, J. M. Effect of male accessory gland products on egg laying in gastropod molluscs. *J Vis Exp* e51698 (2014).

138. Álvarez, B., Koene, J. M. Cognition and Its Shaping Effect on Sexual Conflict: Integrating Biology and Psychology. *Frontiers in Integrative Neuroscience* **16**, (2022).

139. de Vienne, D. M. Lifemap: Exploring the Entire Tree of Life. *PLoS Biol* **14**, e2001624 (2016).

Figures

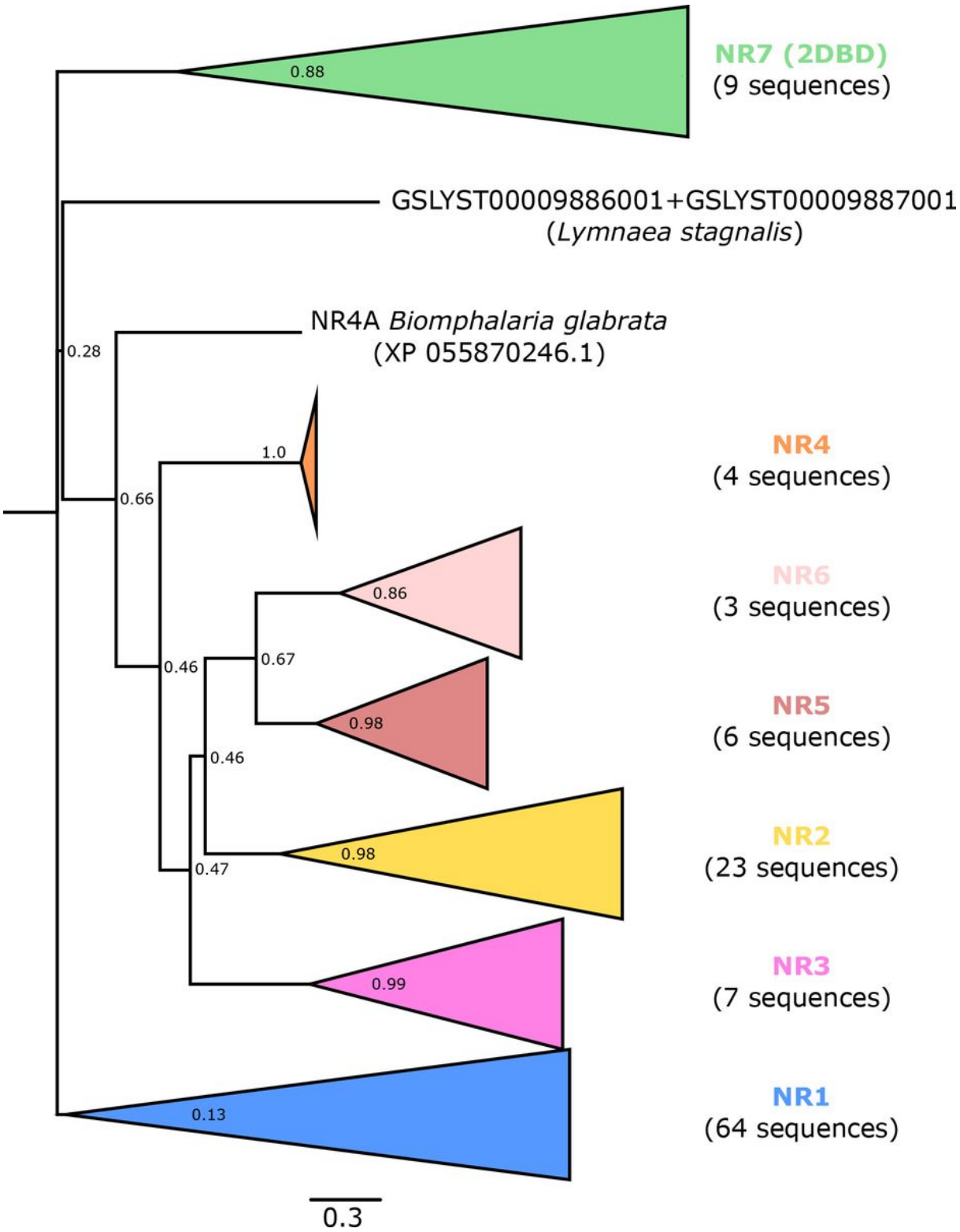

Figure 1

***L. stagnalis* NR relationships**. Phylogenetic relationships among *L. stagnalis* NR families: ML tree (Jones-Taylor-Thornton substitution model, Gamma distributed substitution rate with 4 categories, statistical support from 1,000 bootstraps) based on ClustalW-aligned concatenated sequences of the DBD and most conserved part of the LBD of *L. stagnalis* NRs and their BLAST best-hits (alignment available upon request). Family NR0 was excluded from this analysis as it could not be aligned with DBD-only and LBD-only sequences. The triangles represent collapsed clades, with the number of branches/sequences indicated in brackets (see details on the full tree provided in Supplementary Figure SF2).

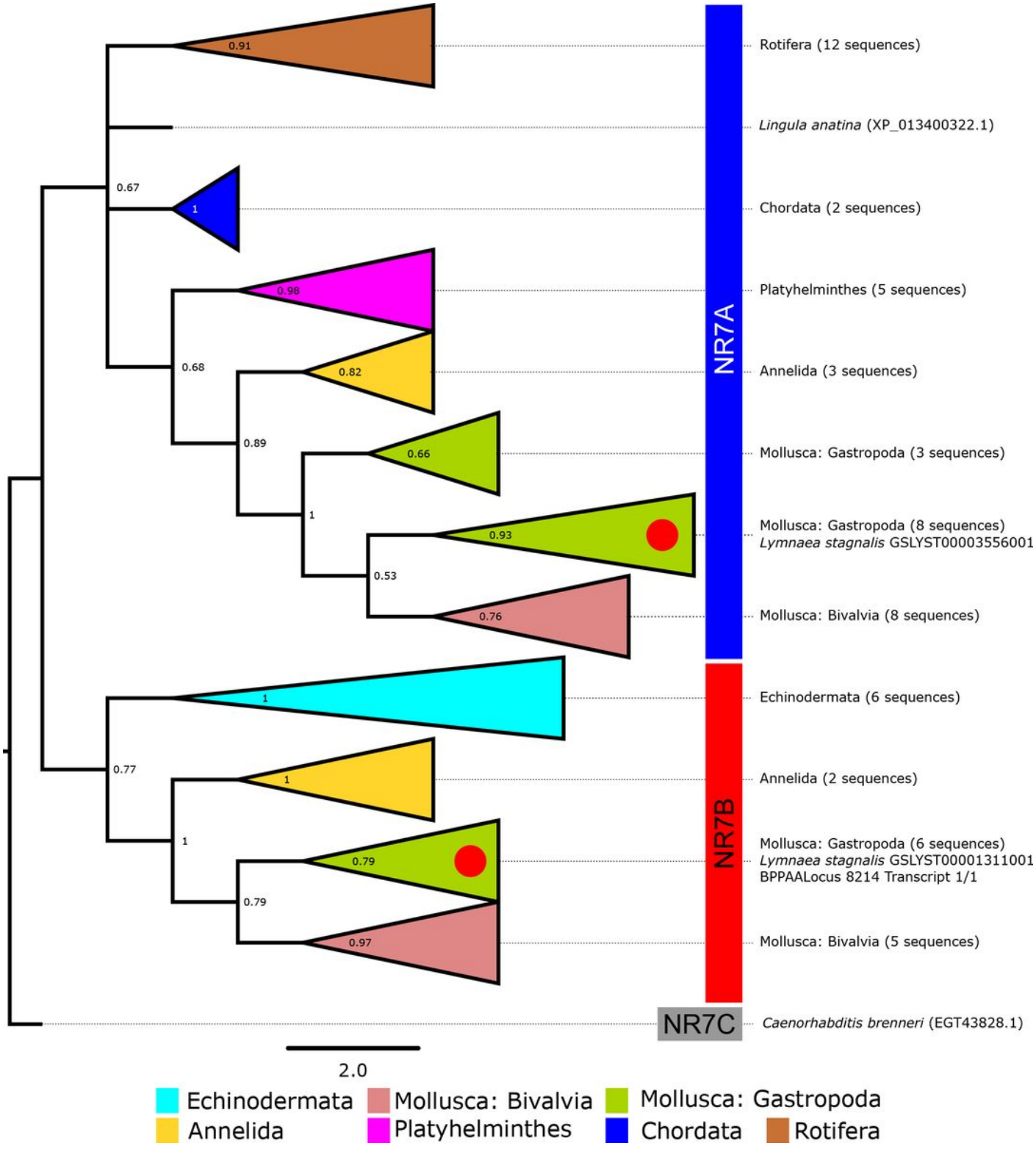

Figure 2

Phylogenetic analysis of 2DBD-NRs relationships across metazoans. Maximum Likelihood (ML) tree (same parameters as in Figure 1; values below branches indicate statistical support based on 1,000 bootstraps) based on the ClustalW-aligned concatenated conserved domains (DBD1, DBD2, and LBD) of the two sequences found in the genome of *L. stagnalis* (present study, indicated by red dots), and 60 sequences identified by Wu and Loverde (2023) from various lineages (Brachiopoda, Rotifera,

Platyhelminthes, Annelida, Mollusca, Nematoda, Echinodermata, Cephalochordata, indicated by the different colours) and having the three domains complete (as verified using Conserved Domains DataBase). The different NR7 groups referred to in the text are indicated on the right in blue, red and grey. Alignment available upon request. See also Supplementary Figure SF3 for detailed clade composition).

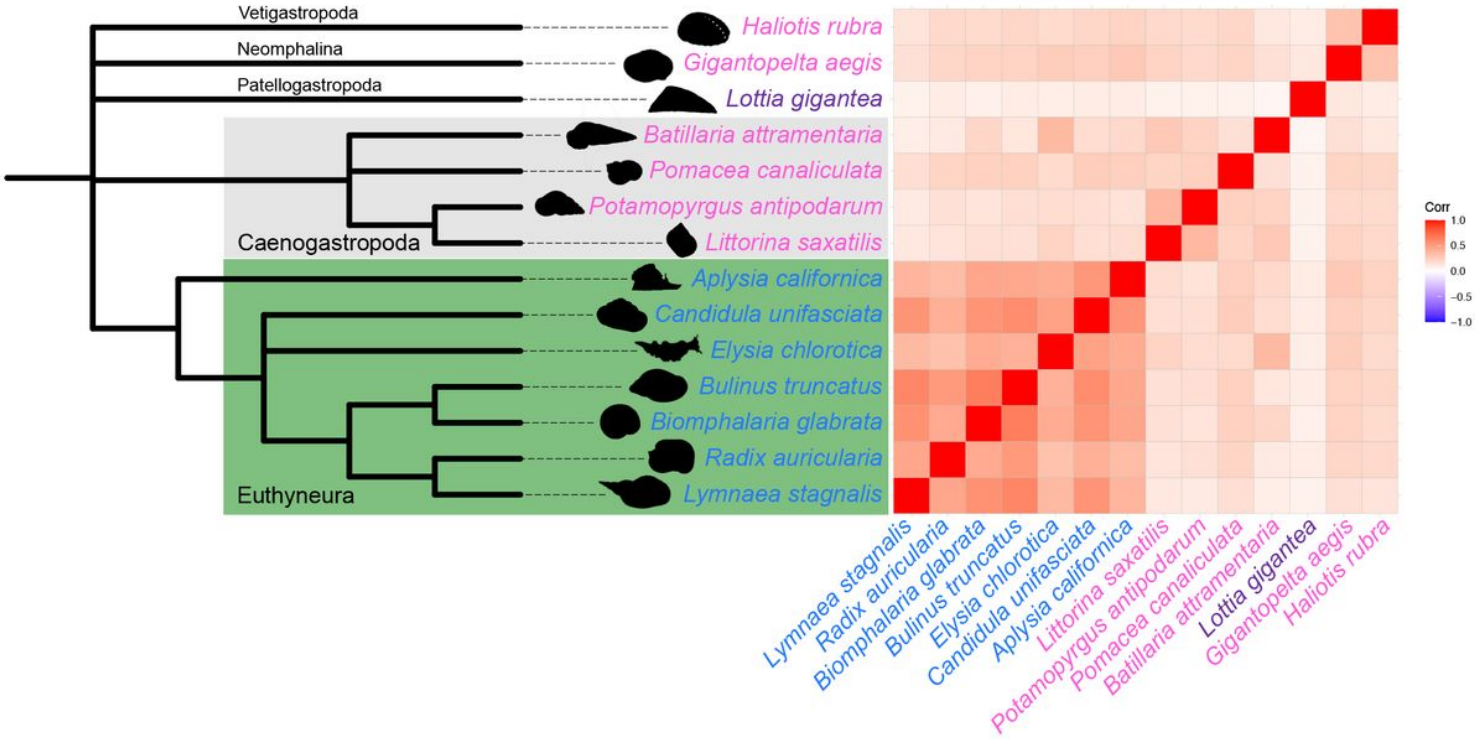

Figure 3

Gastropod proteome orthology. Left: phylogenetic relationships among 14 gastropod species (topology taken from [139]) used for gene orthology assessment (colour code: blue, simultaneous hermaphrodite; purple, sequential hermaphrodite: pink, separate sexed). Right: pairwise correlation of species contribution in terms of number of proteins to 29,428 orthogroups produced by OrthoFinder. Correlations are particularly strong within the set of euthyneuran species used in the analysis. The heat map gradient on the right indicates the level of (positive) correlation.

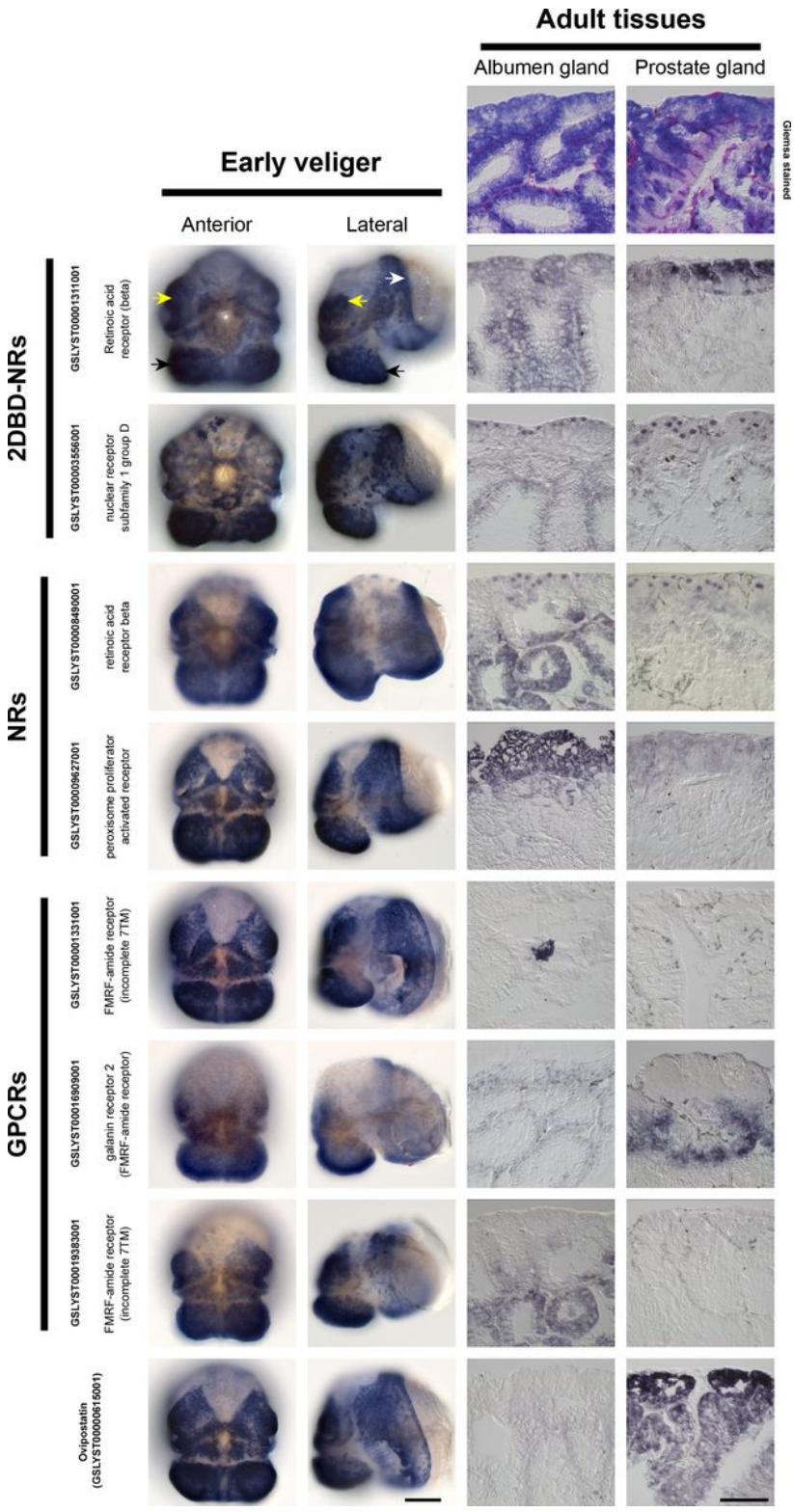

Figure 4

Representative whole mount *in situ* hybridisation (WMISH) expression profiles for a selection of 2DBD-NRs, NRs, GPCRs and ovipostatin differentially expressed between early veliger stage and adult reproductive organs (paired albumen and prostate glands). The expression patterns of these receptors and ovipostatin in early-veliger-stage larvae are broad, and generally similar; all genes are robustly expressed in ectodermal cells covering the foot (black arrowheads), surrounding the stomodaeum (white

asterisk), the mantle (white arrowhead) and the cephalic lobes (yellow arrowheads). In adult tissues ovipostatin is highly expressed in the prostate and effectively undetectable in the albumen gland as expected. Qualitative differences are also evident for many of the receptors surveyed, and hint at possible roles in maintaining these simultaneously active reproductive organs. Representative scale bars are 100 mm.

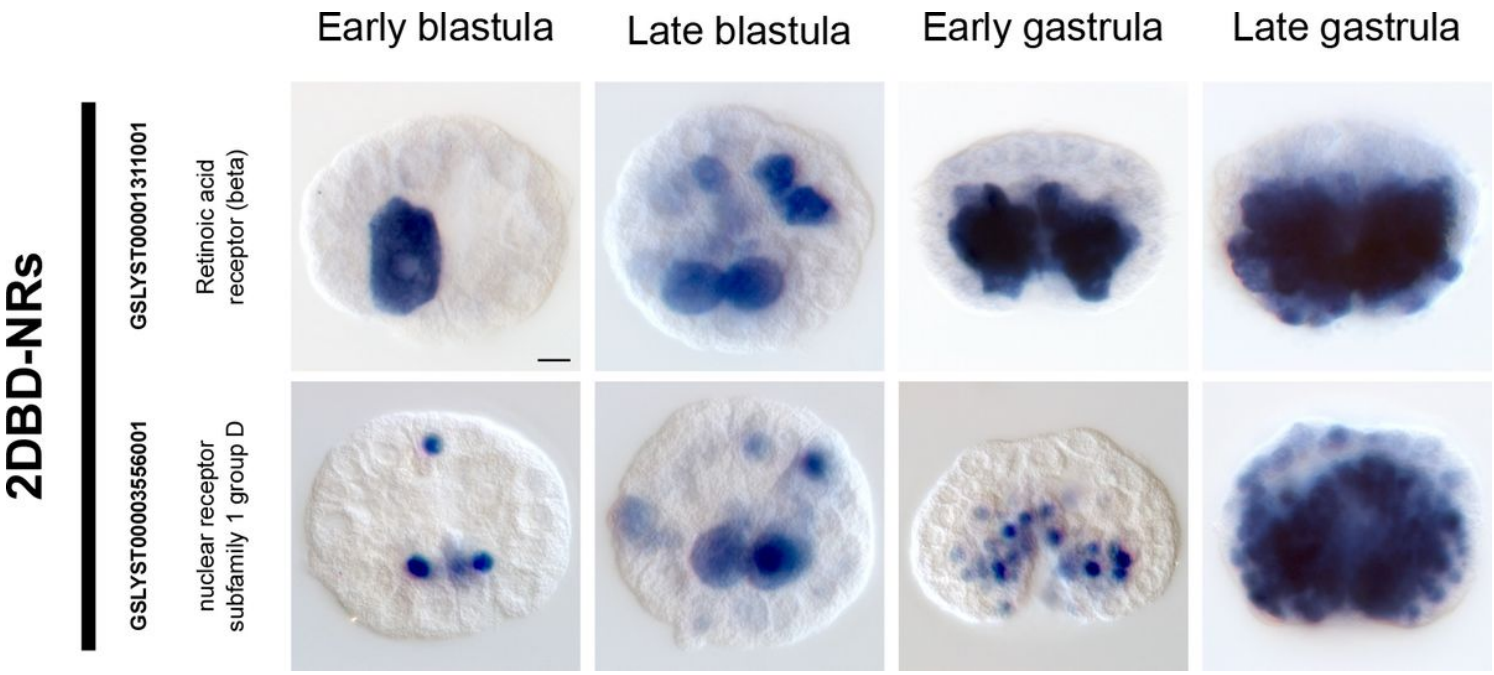

Figure 5

WMISH patterns of 2 2DBD-NRs during early development. Both of these receptors display dynamic, asymmetric patterns of expression prior to and during gastrulation. Expression of GSLYST00003556001 appears to largely be restricted to nuclei, other than 2 large prominent cells in late blastulae. Conversely, transcripts of GSLYST00001311001 are distributed throughout the cytoplasm. Both genes are restricted to the vegetal hemisphere in early gastrulae and are concentrated around the blastopore.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupITFSCIREP.pdf