The Nuclear Receptors of *Biomphalaria glabrata* and *Lottia gigantea*: Implications for Developing New Model Organisms

Satwant Kaur¹, Susan Jobling¹, Catherine S. Jones², Leslie R. Noble², Edwin J. Routledge¹, Anne E. Lockyer¹*

¹ Institute of Environment, Health and Societies, Brunel University London, Uxbridge, United Kingdom, ² School of Biological Sciences, University of Aberdeen, Aberdeen, United Kingdom

* anne.lockyer@brunel.ac.uk

Abstract

Nuclear receptors (NRs) are transcription regulators involved in an array of diverse physiological functions including key roles in endocrine and metabolic function. The aim of this study was to identify nuclear receptors in the fully sequenced genome of the gastropod snail, *Biomphalaria glabrata*, intermediate host for *Schistosoma mansoni* and compare these to known vertebrate NRs, with a view to assessing the snail’s potential as a invertebrate model organism for endocrine function, both as a prospective new test organism and to elucidate the fundamental genetic and mechanistic causes of disease. For comparative purposes, the genome of a second gastropod, the owl limpet, *Lottia gigantea* was also investigated for nuclear receptors. Thirty-nine and thirty-three putative NRs were identified from the *B. glabrata* and *L. gigantea* genomes respectively, based on the presence of a conserved DNA-binding domain and/or ligand-binding domain. Nuclear receptor transcript expression was confirmed and sequences were subjected to a comparative phylogenetic analysis, which demonstrated that these molluscs have representatives of all the major NR subfamilies (1-6). Many of the identified NRs are conserved between vertebrates and invertebrates, however differences exist, most notably, the absence of receptors of Group 3C, which includes some of the vertebrate endocrine hormone targets. The mollusc genomes also contain NR homologues that are present in insects and nematodes but not in vertebrates, such as Group 1J (HR48/DAF12/HR96). The identification of many shared receptors between humans and molluscs indicates the potential for molluscs as model organisms; however the absence of several steroid hormone receptors indicates snail endocrine systems are fundamentally different.

Introduction

The tropical freshwater snail *Biomphalaria glabrata* is an intermediate host for several digenean trematode parasitic worms, including *Schistosoma mansoni*, the causative agent of human
Intestinal schistosomiasis. Human schistosomiasis is the most widespread trematode infection affecting around 200 million people, leading to a chronic debilitating disease and up to 200,000 deaths per year, across 75 developing countries [1]. Due to its medical significance as an intermediate host, B. glabrata has been the focus for much research, including several gene discovery projects [2–4]. Tools have been developed for investigating genomic and transcriptomic attributes of this species, including a BAC library for genome sequencing [5]; a 5K cDNA microarray [6]; a 1.2K oligo microarray [7] and the means to selectively silence gene expression in the snail (RNAi: [8,9]). This background of research has culminated in the sequencing the B. glabrata genome [10] (http://129.24.144.93/blast_bg/index.html). The progress made in developing these resources has also provided the potential for the snail to become a new model organism for other purposes, including testing and identification of endocrine disrupting chemicals (EDCs) and for understanding fundamental biology, including endocrinology.

Endocrine and metabolic disease are among the most common contemporary human afflictions, the prevalence of which has been well defined in large population-based studies (for example, [11,12]). Some of the causes of these are not immediately obvious, but may be related to increasing exposure to EDCs [13]. As a consequence, not only is more testing of potential EDCs needed, but a better understanding of endocrine function and disruption is also required. A solution that is appropriate and in keeping with the three R’s (replacement, refinement and reduction) ethos of animal research [14], is to exploit the use of invertebrate organisms which may offer a simplified model for research, as well as providing a faster, cheaper and more ethically acceptable alternative to mammalian testing, at least during initial chemical screening [15]. As NRs play key roles in endocrine and metabolic functions [16], cross-species comparative studies of the conservation of these genes within invertebrate genomes may identify new model systems for the testing of chemicals with endocrine disrupting potential without using vertebrates. In addition to this, B. glabrata has recently been proposed specifically for developmental toxicity, acute toxicity and mutagenicity testing in order to establish standardised protocols to assess environmental risks [17]. Therefore a better knowledge of normal endocrine function in molluscs will enable us to understand the full impact of EDCs in the environment, where they have been shown to affect both vertebrate and invertebrate species [18].

One class of transcription factor involved in regulating endocrine function in vertebrates are the nuclear receptors (NRs). NRs regulate and coordinate multiple processes by integrating internal and external signals, thereby maintaining homeostasis (reviewed in [16,19]). These proteins exhibit strong similarities in their mode of action due to their conserved molecular structure, which includes a DNA-binding domain (DBD), consisting of two Cys4 zinc fingers, and the ligand-binding domain (LBD), which not only controls signalling by binding small lipophilic molecules, called ligands, but also binds co-activators and co-repressors [20]. Both the LBD and D box domain of the DBD mediate receptor hetero- or homo-dimerization (reviewed in [21]). Biochemical studies and crystal structure of the LBD reveal that ligand binding triggers a conformational change, causing bound co-repressors to be displaced by co-activators, leading to gene expression [22]. NRs are classified into six distinct families by sequence homology using a phylogenetic approach [23]. Based on their conserved nature and their biologically essential roles throughout the Metazoa, NRs are believed to have emerged early in animal evolution, prior to the bilaterian ancestor [24–26].

Molluscs are affected by EDCs and one of the most cited examples is that of the marine pollutant tributyltin (TBT), found in antifouling paint, which is responsible for imposex in at least 195 species of gastropods worldwide (reviewed in [27,28]). TBT has been shown to act through binding to an NR, the retinoid X receptor (RXR) [29,30]. Other environmental pollutants could act through the estrogen receptor (ER) or androgen receptor (AR) and there is also some evidence of a role for androgen and estrogen-like molecules in the reproductive cycle of
molluscs [31,32]. ER orthologues have previously been reported in molluscs (eg [33–37]) although their function is currently unconfirmed [38,36,39,37,40], since, for example, the identified ER homologue from the oyster Crassostrea gigas is unresponsive to estrogen [37]. The presence of an ER in molluscs and the potential to respond to estrogens is confounded by the absence of aromatase (CYP19) [41], the enzyme required for the conversion of testosterone to estradiol, but it is possible that another enzyme catalyses the aromatase reaction [38]. The presence of ARs in molluscs has been inferred [32,42], but this remains an area of controversy [43] since homologues have not been identified despite investigations specifically searching for the gene [44]. A recent survey of NRs in C. gigas also did not identify an AR [45].

Previously only the RXR has been characterised in B. glabrata [46]. With the recent availability of a draft genome for B. glabrata it is particularly timely to search for the NRs of B. glabrata. The genome of the owl limpet, L. gigantea [47], a marine gastropod, for which 26 NRs have already been identified [25,48,49], was compared with the snail NR repertoire. These species, from different families, represent distinct gastropod lineages. The sequences of identified putative genes were confirmed using transcript data from the L. gigantea genome portal and by amplification of products from B. glabrata mRNA. We assess the conservation, divergence, and uniqueness of gastropod NRs in comparison with the previously characterised receptors of a vertebrate (Homo sapiens), insect (Drosophila melanogaster), nematode worm (Caenorhabditis elegans) (reviewed in [50]) and the parasitic trematode (S. mansoni) (reviewed in [51]).

**Methods**

**Identification of nuclear receptors from B. glabrata and L. gigantea genomes**

Human AR (GenBank: EAX05380) and ERα (GenBank: AA128575) gene sequences were used to search for expressed sequence tags (ESTs) in B. glabrata and L. gigantea at National Centre for Biotechnology Information (NCBI) using TBLASTX with default parameters [52]. These ESTs were then used to search for homologues at the preliminary B. glabrata genome (version 4.01: http://129.24.144.93/blast_bg/2index.html) (TBLASTX expect limit:1e-09) and assembled L. gigantea transcripts from the filtered gene models in the JGI Genome Portal (http://genome.jgi-psf.org/Lotgi1/Lotgi1.info.html) using BLASTN (expect limit: 1e-99). This was followed by a wider search using NR gene sequences spanning all the groups present in humans and D. melanogaster. A secondary search using the identified L. gigantea and B. glabrata NR genes against both the gastropod genomes using a low expect threshold value identified additional family members, some of which were distinct from human and fly NRs. Query sequences were filtered for low complexity regions. Identified contigs from B. glabrata were downloaded and processed through GENSCAN [53] (http://genes.mit.edu/GENSCAN.html) using default parameters, to identify predicted coding regions, intron-exon boundaries, and peptides. Predicted peptides were assessed using BLASTP (expect limit: 1e-10) against the non-redundant GenBank database to check for the presence of the NR domains (LBD/DBD). The identified nucleotide sequences from both molluscs containing NR domains were analysed for redundancy using clustering analysis with sequence overlap cut-off set at 0.5 and segment coverage cut-off at 0.25 in Seqtools (8.4ver) (http://www.seqtools.dk/).

The identified genes were named based on the phylogenetic analysis and sequence similarity of the full-length sequence to previously characterized human and D. melanogaster NRs according to the recognised nomenclature [23]. Small sequence fragments of putative NRs from B. glabrata too short to be classified were left out of the subsequent analysis. Where there was no vertebrate or invertebrate homologue, subfamily classification was made, with a number representing the subfamily, a capital letter for the group, and a number for the individual gene.
RNA isolation and cDNA synthesis
Total RNA was isolated from whole homogenized adult *B. glabrata* snails using TRI-reagent (Sigma-Aldrich, St Louis, USA) according to the manufacturer’s protocol and treated with DNase to eliminate contaminating genomic DNA. Total RNA was extracted from embryonic samples using the RNasy Fibrous tissue mini kit (Qiagen, Limberg, Netherlands), which included a DNase treatment as part of the protocol. Quantification and purity of each RNA sample were determined by spectrophotometry (Nanodrop, Thermo Fisher Scientific Ltd. Waltham, USA), and the RNA integrity was visually checked by agarose gel electrophoresis. 4µg of total RNA from adult snails and 1.5µg total RNA from embryonic samples was reverse transcribed in a 20µl reaction using the Superscript III cDNA synthesis kit (Invitrogen, Life Technologies, Carlsbad. USA) with 5µM of a custom oligo (dTAP) primer (TGACTCGAGTCGACATCGAT20) following the manufacturer’s instructions. Residual RNA was removed by adding 1µl of RNase H (2 U/µl) to the reaction and incubating it at 37°C for 20 min. RT-PCR with 18S primers (18S-F: CGCCCGTCTAATTTAC and 18S-R: ACGCCAGCCCAGACCAA) [54] verified successful cDNA synthesis.

Polymerase chain reaction
Specific PCR primers (S1 Table) were designed using PRIMER3 (version 0.4.0: http://bioinfo.ut.ee/primer3-0.4.0/), to amplify fragments from DBD and/or LBD for each *B. glabrata* NR to confirm transcription and sequence. 25µl PCRs contained 2µl cDNA (diluted 1 in 20), 1 X PCR Buffer, 2.5mM MgCl2, 0.2mM dNTPs, 0.5µM forward and reverse primers and 1.25U Taq DNA polymerase (Bioline, London, UK). Reaction conditions were 2 min at 95°C followed by 35 cycles of 30 sec at 95°C, 30 sec at 55–64°C (optimised for each primer pair (S1 Table)) 1 min 30 sec at 72°C, with a final extension of 5 min at 72°C. Amplified products were analysed by gel electrophoresis and products of an appropriate size were gel extracted where necessary and sequenced (Sequencing facility, Wolfson Wellcome Biomedical Laboratory, Natural History Museum, UK). Sequences were deposited in GenBank Accession Nos: JZ390894-JZ390939

Sequence alignments and phylogenetic analysis
The sequenced NR transcripts from *B. glabrata* were translated using proteomics tools at EXPASY (http://www.expasy.org/). The predicted peptide sequences from both *B. glabrata* and *L. gigantea* were analysed using PFAM domain analysis [55] (PF00104 and PF00105) and PANTHER [56], a hidden Markov model-based method (PTHR24082) confirming the NR domains. The NCBI program Simple Modular Architecture Research Tool (SMART, [57]) was used for the identification of DNA-binding domain (DBD) and ligand binding domain (LBD) regions which were then aligned with the DBD and LBD regions of human (*H. sapiens*), fruit fly (*D. melanogaster*) nematode (*C. elegans*) and parasitic trematode (*S. mansoni*) NRs (Table 1). The NR domains were aligned using default parameters in ClustalX and converted to Nexus format using default parameters in Mesquite v2.75. The DBD and LBD from NRs of the other species were obtained using the conserved domain database (CDD) and reconfirmed using SMART. The NR2A subfamily expansion that contains a large number of the *C. elegans* NRs [58] was disregarded and only 15 NRs from *C. elegans* that are broadly conserved among animal phyla were included in this study for comparative analysis.

Phylogenetic reconstruction was performed using Maximum Likelihood analysis with Bayesian inference and MEGA 6.06. Bayesian inference was conducted with MrBayes version 3.1.2 [59] using the WAG model; the best fitting substitution model determined by both Akaike information criterion and Bayesian information criterion frameworks (ProtTest (v1.4) [60]). Two independent runs of Markov Chain Monte Carlo (MCMC) analysis were
### Table 1. Sequence Identification of All Nuclear Receptors for Species in the Study: NRs from Fly, Human, Nematode and Trematode (NCBI Accession Numbers), Compared to Identified *L. gigantea* NRs (Protein Identification Numbers: JGI Genome Portal Version 1.1) and *B. glabrata* NRs (Contig Numbers: Preliminary Bg Genomic Data (Version 4.01)).

<table>
<thead>
<tr>
<th>Group</th>
<th>C. elegans</th>
<th>Accession</th>
<th>S. mansoni</th>
<th>Accession</th>
<th>D. melanogaster</th>
<th>Accession</th>
<th>B. glabrata</th>
<th>Contig</th>
<th>L. gigantea</th>
<th>Protein ID</th>
<th>H. sapiens</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>0A</td>
<td>KNR</td>
<td>CAA31709</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KNRL</td>
<td>AAF51627</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EGON</td>
<td>CAA34626</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0B</td>
<td>BgDAX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LgDAX</td>
<td>153776</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1A</td>
<td>THRα</td>
<td>AAR32912</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>THRβ</td>
<td>AAR29359</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LgDAX</td>
<td>207623</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1B</td>
<td>ODR7</td>
<td>AAC46497</td>
<td>RAR-like*</td>
<td>CCD76558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BgRAR</td>
<td>398</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LgRAR</td>
<td>207867</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1C</td>
<td>BgPPAR1</td>
<td>2052</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LgPPAR1</td>
<td>174409</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BgPPAR2</td>
<td>1275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LgPPAR2</td>
<td>238472</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1D</td>
<td>HR85</td>
<td>AAO39185</td>
<td>E75</td>
<td>AAN11687</td>
<td>BgE75</td>
<td>29</td>
<td>LgE75</td>
<td>136477</td>
<td></td>
<td>Rev-erb-a</td>
<td>AAVH6148</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BgRev-erb</td>
<td>160</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rev-erb-b</td>
<td>AAVH613</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BgNR1D1*</td>
<td>1939</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BgNR1D2*</td>
<td>1958</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BgNR1D3*</td>
<td>201</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1E</td>
<td>E78</td>
<td>AAR30507</td>
<td>E78</td>
<td>AAF51692</td>
<td>BgE78</td>
<td>534</td>
<td>LgE78</td>
<td>163301</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HR23</td>
<td>P41828</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1F</td>
<td>HR3</td>
<td>AAA28461</td>
<td>BgHR3</td>
<td>73</td>
<td>LgHR3</td>
<td>167096</td>
<td></td>
<td></td>
<td></td>
<td>RORα</td>
<td>AAVH831</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BgOR</td>
<td>74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RORβ</td>
<td>AAVH93774</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RORγ</td>
<td>AAVH64751</td>
<td></td>
</tr>
<tr>
<td>1G</td>
<td>HR14</td>
<td>AAA06982</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1H</td>
<td>EcR</td>
<td>NP724456</td>
<td>BgEcR</td>
<td>1481</td>
<td>LgEcR</td>
<td>170342</td>
<td></td>
<td></td>
<td></td>
<td>LXRα</td>
<td>AAV38218</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LXRβ</td>
<td>NP_009052</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FXR</td>
<td>AAIO3574</td>
<td></td>
</tr>
<tr>
<td>1I</td>
<td>VDR</td>
<td>AAB95155</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PXR</td>
<td>AAD05436</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CAR</td>
<td>AAAY56401</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1J</td>
<td>DAF12</td>
<td>AAD34462</td>
<td>HR96(CAR)*</td>
<td>AAV80235</td>
<td>HR96</td>
<td>AAC6928</td>
<td>BgNR1J1</td>
<td>48</td>
<td>LgNR1J1</td>
<td>163618</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HR8</td>
<td>AAP31437</td>
<td>HR96</td>
<td>AAW88541</td>
<td></td>
<td></td>
<td>BgNR1J2</td>
<td>9564</td>
<td>LgNR1J2</td>
<td>63892</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HR48</td>
<td>CAD6502</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BgNR1J3</td>
<td>2148</td>
<td>LgNR1J3</td>
<td>163956</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HR1</td>
<td>AAC48174</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BgNR1J4</td>
<td>3233</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1K</td>
<td>HR1</td>
<td>AAC48174</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2A</td>
<td>HR49</td>
<td>CAD57702</td>
<td>HNF4</td>
<td>CCD80248</td>
<td>HNF4</td>
<td>AAF52702</td>
<td>BgHNF4</td>
<td>2702</td>
<td>LgHNF4</td>
<td>73247</td>
<td>HNF4</td>
<td>NP_849180</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HNF4γ</td>
<td>NP_004124</td>
<td></td>
</tr>
<tr>
<td>2B</td>
<td>RXR1</td>
<td>CCD59553</td>
<td>USP</td>
<td>AAF5707</td>
<td>BgRXR</td>
<td>3695</td>
<td>LgRXR</td>
<td>206562</td>
<td></td>
<td>RXRα</td>
<td>AAI0999</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RXR2</td>
<td>AAD33428</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RXRβ</td>
<td>AAC18599</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RXRγ</td>
<td>AAA80681</td>
<td></td>
</tr>
<tr>
<td>2C/D</td>
<td>HR41</td>
<td>NP_500073</td>
<td>TR</td>
<td>AAV80236</td>
<td>HR78</td>
<td>Q24142</td>
<td>BgTR</td>
<td>2718</td>
<td>LgTR*</td>
<td>231979</td>
<td>TR2</td>
<td>AAA36761</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TR4</td>
<td>NP_003289</td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Group</th>
<th>C. elegans</th>
<th>Accession</th>
<th>S. mansoni</th>
<th>Accession</th>
<th>D. melanogaster</th>
<th>Accession</th>
<th>B. glabrata</th>
<th>Contig</th>
<th>L. gigantea</th>
<th>Protein ID</th>
<th>H. sapiens</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>2E</td>
<td>FAX</td>
<td>AAD55066</td>
<td>TLL</td>
<td>AAW88549</td>
<td>TLL</td>
<td>AAB71371</td>
<td>BgTLX*</td>
<td>3</td>
<td>LgTLX</td>
<td>130367</td>
<td>TLX</td>
<td>AAL05871</td>
</tr>
<tr>
<td></td>
<td>HR67</td>
<td>CAA97428</td>
<td>PNR*</td>
<td>CCD82728</td>
<td>PNR</td>
<td>AAF58145</td>
<td>BgDSF</td>
<td>42</td>
<td>LgDSF</td>
<td>64143</td>
<td>PNR</td>
<td>AAD28301</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DSF*</td>
<td>AAW88537</td>
<td>DSF</td>
<td>AAF52303</td>
<td>BgFAX1</td>
<td>148</td>
<td>LgFAX1</td>
<td>168960</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AAF54133</td>
<td>BgPNR†</td>
<td>3827</td>
<td>LgNR2E</td>
<td>232946</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LgPNR</td>
<td>171552</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2F</td>
<td>UNC55</td>
<td>CAP16277</td>
<td>COUP-TFI*</td>
<td>CCD77973</td>
<td>SVP</td>
<td>NP_524325</td>
<td>BgCOUP-TFa</td>
<td>5703</td>
<td>LgCOUP-TFa</td>
<td>COUP-TFα</td>
<td>AAA04154</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AAW88536</td>
<td></td>
<td></td>
<td>BgCOUP-TFα</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BgER</td>
<td>469</td>
<td>LgER</td>
<td>132166</td>
<td>ERα</td>
<td>AA28575</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ERβ</td>
<td>AAV31779</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>ERR</td>
<td></td>
<td>ERR</td>
<td>AAL37554</td>
<td>BgERR</td>
<td>751</td>
<td>LgERR</td>
<td>168715</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ERRa</td>
<td>AAAH3795</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ERRβ</td>
<td></td>
<td>ERRβ</td>
<td>AAC99409</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ERRγ</td>
<td></td>
<td>ERRγ</td>
<td>EAW93335</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4A</td>
<td>HR6</td>
<td>CAA85271</td>
<td>NR4A5*</td>
<td>AAR28090</td>
<td>DHR38</td>
<td>AAF53914</td>
<td>BgHR38</td>
<td>27</td>
<td>LgHR38</td>
<td>94341</td>
<td>NGFB</td>
<td>CAG32985</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BgNR4a</td>
<td>2842</td>
<td>LgNR4a</td>
<td>182258</td>
<td>NURR1</td>
<td>AAAH9288</td>
</tr>
<tr>
<td>5A</td>
<td>HR25</td>
<td>CAA91028</td>
<td>FTZ-F1α (SF1)*</td>
<td>CCD7387</td>
<td>FTZ-F1</td>
<td>AAN11667</td>
<td>BgFTZ-F1</td>
<td>1318</td>
<td>LgFTZ-F1</td>
<td>196916</td>
<td>SF1</td>
<td>AAAH32501</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BgFTZ-F1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5B</td>
<td>HR39(FTZ-F1)</td>
<td>CDD77680</td>
<td>DHR39</td>
<td>AAN11107</td>
<td>BgHR39</td>
<td>3667</td>
<td>LgHR39</td>
<td>94029</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6A</td>
<td>HR91</td>
<td>CAB60329</td>
<td>DHR4</td>
<td>AAX7355</td>
<td>BgHR4</td>
<td>1866</td>
<td>LgHR4a</td>
<td>128948</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GCNF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LgHR4b</td>
<td>162354</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2DBDNR</td>
<td></td>
<td></td>
<td>2DBDα*</td>
<td>AAW88533</td>
<td></td>
<td></td>
<td>Bg2DBDNR1*</td>
<td>304</td>
<td>Lg2DBDNR1*</td>
<td>168696</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2DBDβ*</td>
<td>AAW88534</td>
<td></td>
<td></td>
<td>Bg2DBDNR2*†</td>
<td>1296</td>
<td>Lg2DBDNR2*†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2DBDγ*</td>
<td>AAW88550</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unclassified</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BgNRU1</td>
<td>221</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BgNRU2</td>
<td>293</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BgNRU3</td>
<td>180</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BgNRU4††</td>
<td>250</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*NRs not used in phylogenetic tree;
† NRs are short fragments that could be multiple partial genes or parts of a full transcript;
‡ NRs not found expressed in *B. glabrata*
performed, with four chains run for 7 million generations with the 'temperature' parameter at 0.10, prior probabilities with default values and sampling every 2000 generations. The first 2 million generations were discarded as burn-in because the Log likelihood values were plotted and found to be asymptotic well before the burn-in fraction. Convergence between the independent MCMC runs was examined by the average deviation of the split frequencies and the potential scale-reduction factor (PSRF), which was 1.00. Clades with posterior probabilities >95% were considered well supported. All PSRF values for MrBayes analyses were 1.00. For Maximum Likelihood analysis, the Jones-Taylor-Thornton (JTT) substitution model \[61\] was used, with a gamma distribution of rates between sites (eight categories, parameter alpha, estimated by the program). Support was evaluated by 1000 bootstrap replicates.

Maximum parsimony was also used to provide additional phylogenetic support for classification, naming, and the phylogenetic relationships observed between \textit{B. glabrata}, \textit{L. gigantea} and nuclear receptors from selected species. Maximum parsimony with heuristic searches, branch swapping set to tree-bisection-reconnection, topological constraints not enforced, and 1000 bootstrap replicates were performed using MEGA6.06 \[62\]. An appropriate out-group to root the sequences was difficult for such a diverse and ancient nuclear receptor family and so the phylogenies were mid-point rooted using Dendroscope (ver3.2.10) \[63\].

**Results and Discussion**

**Nuclear receptor genes**

Thirty nine NR sequences were identified from the genome of \textit{B. glabrata} and a total of 33 from \textit{L. gigantea} consisting of 7 newly identified NRs and confirming 26 sequences which had been previously identified from \textit{L. gigantea} in previous studies [25,48,49] (Table 1). An exhaustive search with mammalian and \textit{D. melanogaster} NRs and with the identified mollusc NRs found no further NR homologs. Representative genes for each of the major nuclear receptor groups were detected, suggesting the assembled genome for \textit{B. glabrata} provides a good representation of total gene content with physical genome coverage of approximately 27.5X (Genbank: APKA00000000.1). The numbers found are comparable to the 48 NRs reported in the human genome [64], 49 in mice [65] and 47 in rats [66]. Insects have lower numbers; only 21 NR genes are found in \textit{D. melanogaster} [67], 22 in \textit{Apis mellifera} [68], and 21 in \textit{Tribolium castaneum} [69]. Over 270 NRs have been found in \textit{C. elegans} [70] but only 21 NRs in the nematode \textit{S. mansoni} [71]. The unusually large number of NRs in \textit{C. elegans} is due to extensive proliferation of one gene (HNF4) within the nematode phylum [58]. 43 NRs were recently identified from the Pacific oyster, \textit{C. gigas}, a marine bivalve, [45], similar numbers to those we identified from the gastropods.

**Expression of \textit{B. glabrata} nuclear receptor mRNAs**

Each predicted nuclear receptor was evaluated for expression in \textit{B. glabrata} adult snails, by amplifying cDNA fragments with receptor-specific primers, to demonstrate that the predicted genes were expressed as transcripts and confirm their sequence. Primer pairs for which fragments were not obtained from adults were also tested on embryos. Expression was confirmed for 34 out of 39 identified receptor genes (Table 1), 31 from cDNA derived from whole adult snails, while 3 nuclear receptors, BgTLX, BgDSF and BgFAX1, were not expressed in adult snails but were identified in 96hrs/120hrs embryos. These receptors belong to NR2E sub family and may be involved in embryonic development as reported in \textit{Daphnia pulex} [72]. All amplicons corresponded to the predicted sequence length and sequencing confirmed the predicted identity of the gene products (GenBank Accession Nos: JZ390894-JZ390939). For \textit{L. gigantea}, \textit{in silico} searches identified sequenced transcripts for 21 out of 33 NR genes (S2 Table). Three
B. glabrata putative NRs (BgNR1D1/2/3) for which only partial sequences were obtained from the genome were not included for further phylogenetic analysis, since it was not possible to ascertain if they are multiple partial genes or parts of the same transcript. In addition to this BgTLX, BgNRU4 and LgTR were not analysed further as these were also incomplete sequences. A separate analysis was made of 2DBD genes from B. glabrata and L. gigantea with other previously identified 2DBD sequences from S. mansoni and other species as the 2DBD region could not be aligned with the other NRs.

Phylogenetic analyses

The DBD and LBD regions for the full-length molluscan NRs (32 for B. glabrata and 30 for L. gigantea) were aligned with S. mansoni, D. melanogaster, C. elegans and H. sapiens NR sequences (S1 File). Phylogenetic trees were constructed using 3 different approaches: Bayesian Inference (BI), Maximum Parsimony (MP) and Maximum Likelihood (ML) and the resulting trees assessed for concordance. All trees agree at family and subfamily level (S1 Fig, S2 Fig, S3 Fig) but BI shows the greatest resolution at the base of the phylogram and is shown with nodal support generated from all the tree construction methods (Fig 1). The tree shows the main groupings of the NRs.

The position of the unclassified B. glabrata NRs (BgNRU) is the same in all trees, close to NR0A subfamily within the NR1 group; however the NR0 subfamily was derived to accommodate nuclear receptors that lack either the DBD or the LBD and these 3 B. glabrata NRs contain both domains (S4 Fig). Therefore these remain as unclassified. The position of members of NR2C/D is also not resolved within Group 2 (Fig 1). Both mollusc genomes each contain one member of this clade that we have designated LgTR/BgTR based on their sequence similarity with HsTR (Table 1). LgTR was not included in the phylogenetic analysis since it was missing the LBD, but BgTR clusters with the other NR2C/D group members Ce_HR41, SmTR and Dm_HR78. The Bayesian tree predicts BgTR as the most distal member of NR2C but the MP and ML trees show both the mollusc sequences to be closely related to DmHR78. However the MP tree placed Ce_HR41 in NR2D, although this NR is usually classified in NR2C. Clearly, the NR2C/D groups seem to be related and not well defined. Within NR1 NR1D was not clearly resolved. In the NR3 subfamily, only the position of the human GR showed non-concordance between the BI tree and the MP tree. All of the trees were in agreement for the NR4/NR5/NR6 subfamilies, with one difference in the position of BgNR4a/LgNR4a and NR5A in the MP tree.

The identified genes have been named in agreement with the unified nomenclature for NRs [23] based on their sequence homology and phylogenetic position. Overall, the identified mollusc NRs encompass almost all the subfamilies, with representatives divided between 21 groups (Table 1) and revealed both similarities and differences in the NR complement between the species used in the study. Seventeen of the 19 subgroups containing human NRs also had representatives in the invertebrate species, including molluscs although in general H. sapiens contained several NRs where snails had only one, most likely the result of gene or lineage-specific duplication events. The presence of these groups among both protostomes and deuterostomes suggests that these receptors originated in a common ancestor of the bilateria [24,25]. The 2 vertebrate groups not represented in molluscs were the Groups 3C and 1I. Three groups that were common to humans and molluscs, but not found in the other invertebrate organisms included here were 0B (DAX), 1C (PPAR) and 3A (ER). The presence of these receptors in molluscs suggests at least the possibility of commonality in shared signalling pathways, elucidating the evolutionary development of the endocrine system. Seven groups were found only in invertebrates; Groups 1G and 1K only in C. elegans; 0A only in D. melanogaster, while groups 1E (E78), 5B (DHR39) 1J and the 2DBD NRs were found in several of the invertebrate species.
NRs from six species *B. glabrata* (Bg), *L. gigantea* (Lg), *H. sapiens* (Hs), *D. melanogaster* (Dm), *C. elegans* (Ce) and *S. mansoni* (Sm), were subjected to phylogenetic comparisons using Bayesian inference, maximum parsimony and maximum likelihood methods. The Bayesian tree (midpoint rooted) is shown with posterior probability values from Bayesian inference and bootstrap values from maximum parsimony and maximum likelihood trees. The value of 1 on each node represents 100% posterior probability/bootstrap support; an X indicates an area of disagreement from the Bayesian tree (S1 Fig). Scale bar, 0.1 expected changes/site.

doi:10.1371/journal.pone.0121259.g001
Representation of these latter groups in *L. gigantea* and *B. glabrata* shows that these nuclear receptors pre-date the ecdysozoan/lophotrochozoan split.

**Group 3 receptors**

The nuclear receptors that bind to steroids, such as androgens, estrogens and progestogens, are responsible for the long-term effects of steroid hormones on reproduction, behaviour, immunity, stress responses and development. The Group 3 steroid receptor family includes the estrogen related receptor (ERR), as well as the estrogen receptor (ER) and group 3C subfamily with the androgen, progesterone, glucocorticoid and mineralocorticoid receptors. Steroid receptors were originally thought to be a vertebrate specific gene family but the identification of the genes with a clear homology to human ERR in trichoplax [73] and an ancestral steroid receptor in amphioxus [74] suggested that these receptors arose early in metazoan evolution and subsequently proliferated in vertebrates through series of gene duplication events [75]. We have identified ER homologues in *B. glabrata* and *L. gigantea* (Table 1, Group 3A), which have been previously reported in other molluscan species, including bivalves, *C. gigas* [37], *Mytilus edulis* [76] and *Chlamys farreri* [77] and gastropods, *Aplysia californica* [38], *Lymnaea ollula* [78], *Marisa cornarietis* [33], *Thais clavigera* [39], *Nucella lapillus* [79], *Bithynia tentaculata* [35] and the cephalopod, *Octopus vulgaris* [36]. The phylogenetic position of the mollusc ER homologue with the vertebrate ER is well supported with a BMCMC posterior probability of 98% and bootstrap values of 0.88 and 0.79 (Fig 1), supporting the suggestion that vertebrate and invertebrate ER diverged from a common ancestor, before the evolution of the deuterostomes [38,75,80]. Structural similarities between ER DBDs of molluscs, annelids, cephalochordates and vertebrates have been documented showing them to bind to and regulate transcription through estrogen response elements [36,40,74,81,82]. The ER LBD in invertebrates may have unique functions, since the mollusc ER homologue, at least in vitro (usually in mammalian reporter cell lines), appears to activate transcription constitutively in the absence of a ligand [36,37,39,40] although in annelids the ER has been shown to activate transcription in the presence of estrogens [82]. We also identified a single estrogen-related receptor (ERR) in both *B. glabrata* and *L. gigantea* (Table 1, Group 3B), which clusters with fly and human ERRs. ERRs have been reported previously in molluscs, including *M. cornarietis* [33], *C. gigas* (Genbank: EKC20050), *Mizuhopecten yessoensis* (GenBank: BAN84542.1) and *M. edulis* [76]. ERRs have no known endogenous ligand, although they are thought to bind to estrogen response elements and may play a role in estrogen signalling and energy metabolism [83].

We identified no convincing homologues in either of the two molluscs for the AR or for any other members of group 3C (Fig 1), which also contains the glucocorticoid receptor (GR), mineralocorticoid receptor (MR) and progesterone receptor (PR). The absence of an AR sequence homologue, both in our systematic exhaustive genomic searches of 2 gastropod species, as well as specific laboratory-based searches in other molluscs [44] using an approach which successfully identified other receptors and another systematic NR survey in the bivalve *C. gigas* [45] strongly suggests no vertebrate AR homologue exists in gastropod molluscs. The presence of an AR in molluscs has been hotly contested [84], its existence having been inferred from the effects of androgens and anti-androgens in several species of mollusc (eg. [42,85]), rather than by the identification of a homologous sequence. The absence of an AR means that the findings from previous papers requiring its presence to explain their results may require further analysis in the light of this information; although it should also be noted that the effects of steroids, including those reported for androgens in molluscs may not be mediated via NRs. Steroid hormones have also been shown to act via non-genomic mechanisms in vertebrates, using
membrane bound receptors from the G protein family [86–88], although homologues for these receptors have also not yet been identified in molluscs.

The absence of a molluscan AR and the constitutive expression of the ER in vitro suggest alternative pathways must exist for spermatogenesis/oogenesis in molluscs and other nuclear receptors have been proposed as initiating these pathways for reproductive processes, some of which exist in vertebrates and which may be particularly important in invertebrates. Although vertebrates have subsequently evolved further processes involving steroid hormones, the presence of orthologous NRs involved in other pathways in molluscs offers the potential opportunity to study these conserved networks in an invertebrate.

### Group 1 and 2 nuclear receptors

Conservation between other NR-mediated pathways in molluscs and humans, even with the absence of Group 3C NRs, still provides the potential to develop a mollusc model for endocrine physiology. The retinoid X receptor (RXR, Group 2B) has been previously identified in molluscs including *B. glabrata*, [46], *L. gigantea* [25] and *L. stagnalis* [89]. We found that the molluscan RXR sequences grouped with good support (BMCMC posterior probability of 91%) close to the human RXR sequences, with the *D. melanogaster ultraspiracle* (USP) RXR homologue basal to the clade (Fig 1). BgRXR has been demonstrated to bind and be activated by retinoids, suggesting that retinoid signalling pathways are conserved in the Lophotrochozoa [46]. It was also shown to bind to the response element DR1 either as a homodimer or as a heterodimer with mammalian RARα, LXR, FXR or PPARα [46], which suggests the possibility of conservation of several important signalling pathways in molluscs. The ligand and the co-activator peptide have been previously reported to bind to snail RXR in essentially the same manner as observed in human RXR LBD structures [90], suggesting that the mechanisms of RXR-mediated transcription regulation are very similar in molluscs and humans. The significance of RXR signalling on reproductive physiology in molluscs is strongly supported by the finding that RXR expression is affected during TBT exposure and subsequent development of imposex in *N. lapillus* [91]; RXR is thought to be the receptor which mediates TBT induction of imposex in this species [29]. It has also been recently demonstrated that, in conjunction with RXR signalling, peroxisome proliferator-activated receptor (PPAR) pathways are activated by organotins [92] and may induce imposex in response to TBT in *N. lapillus* [93]. Convergence of 9-cis retinoic acid, a natural ligand for RXR, and PPAR signalling pathways through PPAR-RXRα hetero-dimerization is well established in mammals [94]. We identified two PPAR homologues in both *L. gigantea* and *B. glabrata* both containing identical sequence to the P-box of human PPAR (CEGCKGFFRRTI) which grouped with the human PPAR genes (Fig 1); however the 3 human genes seem to have diverged after the split with the molluscan paralogues. Examination of PPAR gene relationships in more detail and including some additional species in the phylogenetic analysis, confirms that the 2 molluscan paralogues diverged prior to the split with vertebrates. The genes we have designated PPAR1 lie basal to the clade, while the vertebrate α, β and γ PPAR genes are orthologues of the gene we have designated PPAR2 (Fig 2).

The RXR also binds with the thyroid hormone receptor (THR, Group 1A). Both gastropods contain a THR homologue (Table 1), which has been previously reported in *L. gigantea* and other invertebrates [49]. Wu et al (2007) confirmed that the *S. mansoni* THR was able to heterodimerise with the RXR. The molluscan THRs cluster with the human THRα/β (Fig 1). The presence of this NR makes *B. glabrata* a potential model for thyroid hormone processes. Thyroid hormones (THs) play key roles in growth development and metabolism in vertebrates. In invertebrates exogenous (food) sourced THs have been suggested to be involved in signalling [95].
Ultraspiracle (USP), the fly orthologue of the RXR (reviewed in [96]), has been shown to heterodimerize with the ecdysone receptor (EcR) to control ecdysone signalling in insects driving metamorphosis and moulting [97]. EcR homologues (Group 1H) were identified in both molluscs (Table 1), the first time this has been reported for *B. glabrata*. The presence of the EcR ligand, ecdysone, has been previously reported in pulmonate snails, in *B. glabrata* and *Lymnaea stagnalis* [98], but its origin (endogenous or exogenous) is still contested [99]. Interestingly, ecdysone exposure was found to affect growth and egg production in *B. glabrata* and it has been suggested that ecdysone secreted from the *S. mansoni* parasite might provide a
mechanism for parasitic castration and gigantism seen in some intermediate hosts when infected [100]. The EcR of *L. gigantea* has been previously identified and its 3D structure investigated [101] revealing the ligand-binding pocket of the *L. gigantea* EcR homolog has the potential to bind to ecdysone-related steroids. The observation that EcR is expressed in testis tissue may also indicate a role in molluscan reproductive processes [101].

One retinoic acid receptor (RAR) was identified in each gastropod species. RARs have previously been identified in the gastropod *L. stagnalis* (GenBank: GU932671), the rock shell, *T. clavigera* [102], the bivalve, *C. gigas* [45] and recently in *N. lapillus* [103]. In vertebrates RAR binds retinoic acid, the biologically active form of vitamin A, which mediates cellular signalling in embryogenic antero-posterior patterning of the central nervous system [104]. However the RAR of *N. lapillus* did not bind to all-trans retinoic acid or any other retinoid tested nor was it able to activate the transcription of reporter genes in response to stimulation by retinoids [103]. More work is clearly necessary to elucidate the function of RAR in molluscs.

We also identified homologues of Group 2 C/D vertebrate testicular receptor 4 (TR4) in *B. glabrata* and *L. gigantea*, (designated as BgTR and LgTR) that in mice have been shown to control spermatogenesis [105] and folliculogenesis [106]. As an orphan nuclear receptor (with no known ligand) the physiological function of TR4 has been difficult to ascertain until the recent development of knock-out mice for this gene (reviewed in [107]). Thus determining the basic biology of vertebrate NRs is dependent on animal experimentation, some of which may be possible in a simplified model system such as a mollusc.

The NR2E nuclear receptors that have been functionally characterized have a common role in nervous system development, for example the *tailless* (*tll*) gene of *D. melanogaster* is involved in embryonic CNS and larval eye development [108] and the mouse *Tlx* gene has also been found to be a key component of retinal development and vision [109]. In *B. glabrata* we identified homologues of photoreceptor cell-specific nuclear receptor (PNR), *dissatisfaction* (DSF), TLX and FAX1, while in *L. gigantea*, we identified a further NR2E gene (Table 1).

Groups NR1J and NR1I cluster in the tree (Fig 1). No members of the NR1I group, including vitamin D receptor (VDR) or pregnane X receptors (PXR), were identified in *B. glabrata* and *L. gigantea*; however NR1J group in protostomes shares similarity with vertebrate NR1I group [110] and there is evidence that both NR1I/NR1J groups share a common ancestor [24]. NR1J receptors are considered as natural sensors and are involved in xenobiotic metabolism in vertebrates [111]; however, other studies have indicated that NR1J members might regulate xenobiotic responses in *D. melanogaster* and *C. elegans* [112,113]. Our results show that *L. gigantea* has three and *B. glabrata* four NR1J representatives (Table 1). The mollusc NR1J group members possess the well-conserved base contact residues (ESCKAFFR) within the DBD, characteristic of the NR1J sub family [114]. In molluscs the NR1J group receptors may be able to perform the same xenobiotic recognition functions as the closely related NR1I of vertebrates.

**Group 5 and 6 nuclear receptors**

Germ cell nuclear factor (GCNF) homologues (Group 6) have previously been reported in bilaterians (reviewed in [50]) and we identified them in both gastropod species. This is in contrast to the recently published survey of NRs in the pacific oyster *C. gigas* [45] which did not identify any NRs from Group 6. Both BgHR4 and LgHR4a/b cluster with good support (posterior probability 1) to the *D. melanogaster* GCNF homologue, *Dm_HR4* (Fig 1). During embryonic stages in vertebrates, GCNF can interfere with retinoic acid signalling affecting the expression of *cyp26A1*, which is essential for normal hindbrain patterning and early developmental stages [104]. In Group 5 we identified BgFTZ-F1/LgFTZ-F1 and BgHR39/LgHR39, homologues of the *D. melanogaster* NR5 subfamily members, *fushi tarazu* factor 1 and hormone receptor 39
respectively, with a conserved stretch of 23 amino acids adjoining the C-terminal end of the zinc finger motif (AVRSDMRGGRNKFPGPMYKRDRA). This sequence is located immediately after the DBD and plays an important role in high affinity interactions of the receptor with DNA [115]. These Group 5 NRs are related to the vertebrate SF-1 which controls reproductive development and regulates the transcription of steroid-modifying cytochrome P450 genes. In *Drosophila* HR39 is essential for sexual development, required in females both to activate spermathecal secretion and repress male-specific courtship genes, and controls the expression of specific cytochrome P450 genes [116]. The conservation of function between invertebrates and vertebrates in these receptors in Groups NR5 and NR6, suggests the potential for molluscs to be used to model some aspects of mammalian reproductive biology and that further study into the reproductive biology of invertebrates is warranted.

2DBD nuclear receptors

We identified two 2DBD-NRs in *B. glabrata* as well as the two 2DBD-NRs previously reported from *L. gigantea* [48]. We compared these to the 3 previously identified from *S. mansonii* [71] as well as 2DBD NRs from other organisms (Fig 3). The gastropod 2DBD1 clusters with Sm2DBDγ while the position of the second gastropod 2DBD was not determined within the clade. Both the identified Bg 2DBD-NRs possess the two tandem DNA binding domains, and a well-conserved P-box sequence (CEACKK) in the first DBD and a P-box (CEGCK) in the second DBD. The P-box of the second DBD is similar to DBD of NR1 family but the P-box of the first DBD is unique to this NR and may determine novel targets [48].

NRs in parasite and intermediate host

The discovery of differences and similarities in the nuclear receptor repertoire of snail host and *S. mansonii* may open up avenues to further characterize host-parasite interactions and potentially to interfere with schistosome development within the host. Parasites are known to interfere with the snail reproductive system, since one effect of trematode infection in snails is parasitic castration (reviewed in [117]). The intimate host-parasite and intermediate host-parasite relationships, makes a comparison of nuclear receptors in all three organisms significant. It has been suggested, at least for the definitive host, that the ability of the parasite to exploit the hosts’ hormonal microenvironment may be critical to allow it to establish, grow and reproduce [118]. This may also be significant in enabling the parasite to inhabit its intermediate host and may depend upon shared nuclear receptors. All 3 organisms have one or more THR, RAR, HNF4, RXR, TR, COUP-TF and similar NRs in Groups 2E, 4A and 5A. Only snails and schistosomes have the E78, HR39, HR96 and 2DBD NRs, while snails and humans only have DAX1, PPAR, Rev-erb, ROR, ER, ERR and group 6A NRs. NRs which are specific to the snails (in this three-way comparison) are the EcR and the group of unclassified NRs, designated as BgNRU1, BgNRU2, BgNRU3, and BgNRU4 respectively and placed in a separate group (Table 1). These unclassified receptors may have originated from a specific duplication event in a *B. glabrata* precursor, or alternatively, they could be remnants of ancient NR subfamilies, whose representatives have been secondarily lost in the other represented species. With regard to the EcR, given that snails have a EcR homologue and schistosomes do not, the suggestion that β-ecdysterone acts as an attractant in host location by miracidia and affects the rate of growth and maturation in snails [100] is potentially interesting, although the ability of snails to make ecdysone is not proven [99]. Both organisms have homologues of the *D. melanogaster* NRs E78 and HR39 which are active in ecdysone signalling [116,119].

In addition to discovering the intricacies of the molecular interactions between host and parasite with a view to disrupting the development of the schistosome, the other key strategy
for schistosomiasis control is to reduce the numbers and/or distribution of the intermediate host snails. Currently molluscicides, such as the commercially available niclosamide-based Bayluscide, are effectively employed in schistosomiasis control programs; however the problems of resistance and toxicity to other organisms means that the search for alternative, more selective, compounds is on-going. The capacity of nuclear receptors to bind small ligands, including exogenous substances such as natural products and synthetic chemicals, makes them potential targets for molluscicides [120].
**Biomphalaria glabrata as a model organism**

There is a long tradition in biology of examining biochemical processes in simplified models to elucidate mechanisms in more complex organisms. We have demonstrated that there are some NR groups with a single representative in the mollusc species examined instead of the multiple genes from humans. For example in the NR1 subfamily, both molluscs have single homologues of several significant NRs, such as the retinoic acid receptor (RAR), THR and ROR. For these groups, *B. glabrata* could provide a simple model system, not only to study the development and diversification of endocrine systems, but also to investigate and test gene function and response to external stimuli. The identification of many vertebrate-like NRs in *B. glabrata* could make it a suitable model candidate to investigate functional relationships of individual receptors. The ultimate choice of model organisms lies not only with the biology of the organism, but also on its tractability for study and manipulation. Aside from the genome, the full range of genomic tools available for *B. glabrata*, such as RNAi [8,121], BAC library [5] and microarrays [6,122,123] as well as the potential to use classical approaches of enhancer and suppressor genetics and transgenics to explore regulatory networks, make *B. glabrata* a good candidate. However, the advent of next generation sequencing makes transcriptome study from non-model organisms a real possibility and therefore other mollusc species may ultimately make better model systems. In particular, this approach could be used to identify inputs from signal transduction pathways, potential hormone metabolic genes, co-activators, co-repressors and other unknown factors that may impinge upon receptor activity. Based on the absence of Group3C NRs, it can be concluded that *B. glabrata* is an inappropriate model for mammalian steroid hormone function mediated via NRs, as the genes for several major steroid hormone receptors are not present. However since steroid hormones have also been shown to act via non-genomic mechanisms in vertebrates, using membrane bound receptors (mPR and GPR30) [87] there is still potential to examine alternative pathways in these organisms.

**Conclusions**

We identified 39 nuclear receptor genes in *B. glabrata* and 33 in *L. gigantea* representing all seven principal vertebrate nuclear receptor groups. Molluscan endocrinology is poorly understood, so further study to determine the functionality of these identified NRs promises new insights, especially concerning the many unanswered questions regarding the effects of steroids and other EDCs on molluscs. The importance of the snail as an intermediate host for schistosomes, also justifies further investigation into the function of such genes. The absence of vertebrate NRs such as VDR, CAR and PXR, as well as steroid hormone NRs, AR/PR/GR/MR for which we found no orthologues in the two mollusc species examined in this study, indicates that several significant signalling pathways are absent in gastropods. Nevertheless, we have identified an array of NRs common to both vertebrates/molluscs and molluscs/flies. These results add weight to Thummel’s speculation about the convergent regulation of NRs in vertebrates and invertebrates [124]. Elucidation of NR targets in molluscs may unlock their potential as new model organisms allowing the discovery of new pathways leading to similar phenotypes found in vertebrates or, indeed, similar pathways that produce a different phenotype; both of which could potentially form simple test assays, for determining gene function or drug/chemical testing. The range of phenotypes, and their underlying genetic mechanisms, available for study in different species may enable the identification of alternative pathways mediated by NRs that might also be exploited [125]. The potential of many invertebrate species for endocrine study is yet to be explored, but the underlying fundamental similarities and differences between molluscs and vertebrates may be the solution to determining not only the endocrine mechanisms of molluscs but also the full intricacies of our own.
Supporting Information

S1 Fig. Bayesian tree showing nuclear receptors in molluscs, human, fly, nematode and trematode. The NRs from six different species were subjected to phylogenetic comparisons using Bayesian inference. The Bayesian tree (midpoint rooted) with posterior probability values is shown. Notations Bg, Lg, Hs, Dm, Ce and Sm in association with receptor names denote sequences from B. glabrata, L. gigantea, H. sapiens, D. melanogaster, C. elegans and S. mansoni respectively. (PDF)

S2 Fig. Maximum Likelihood tree showing nuclear receptors in molluscs, human, fly, nematode and trematode. The nuclear receptors from six different species were subjected to phylogenetic comparison using maximum likelihood method using Jones-Taylor-Thornton (JTT) substitution model. Node labels indicate bootstrap values. Notations Bg, Lg, Hs, Dm, Sm and Ce in association with the receptor name denote sequences from B. glabrata, L. gigantea, H. sapiens, D. melanogaster, S. mansoni and C. elegans, respectively. (PDF)

S3 Fig. Maximum Parsimony tree showing nuclear receptors in molluscs, human, fly, nematode and trematode. The nuclear receptors from six different species were subjected to phylogenetic comparison using maximum parsimony. Node labels indicate bootstrap values. Notations Bg, Lg, Hs, Dm, Sm and Ce in association with the receptor name denote sequences from B. glabrata, L. gigantea, H. sapiens, D. melanogaster, S. mansoni and C. elegans, respectively. (PDF)

S4 Fig. Sequence alignment file showing the unclassified NRs from B. glabrata as different genes with overlapping regions (DBD and LBD). (PDF)

S1 File. A clustal alignment file of the conserved domain regions (DBD and LBD) of all the nuclear receptors included in the study. Notations Bg, Lg, Hs, Dm, Ce and Sm in association with receptor names denote sequences from B. glabrata, L. gigantea, H. sapiens, D. melanogaster, C. elegans, and S. mansoni respectively. (TXT)

S1 Table. Oligonucleotide primers used in RT-PCR analyses of nuclear receptor genes. Forward and reverse primer sequences with their product size (as submitted to GenBank), their corresponding GenBank accession numbers and their optimised annealing temperatures. (DOCX)

S2 Table. L. gigantea transcripts. Detailed summary of all the top BLAST expressed sequence tag (EST) hits for L. gigantea NRs, including the EST description, GenBank accession number and E-value. (PDF)

Acknowledgments

We thank the B. glabrata Genome Initiative steering committee for their approval to publish this analysis from the unannotated B. glabrata genome (version 4.01).
Author Contributions
Conceived and designed the experiments: SK SJ CSJ LRN EJR AEL. Performed the experiments: SK AEL. Analyzed the data: SK AJ LRN EJR AEL. Contributed reagents/materials/analysis tools: SK SJ CSJ LRN EJR AEL. Conceived and designed the experiments: SK SJ CSJ LRN EJR AEL. Performed the experiments: SK AEL. Analyzed the data: SK AEL CSJ. Contributed reagents/materials/analysis tools: SK AEL CSJ. Wrote the paper: SK AJ LRN EJR AEL.

References


48. Wu W, Niles EG, Hirai H, LoVerde PT. Evolution of a novel subfamily of nuclear receptors with mem-

49. Wu W, Niles EG, Loverde PT. Thyroid hormone receptor orthologues from invertebrate species that each contain two DNA binding domains. BMC Evol Biol. 2007; 7: 27. doi: 10.1186/1471-2148-7-27 PMID: 17319933


