Journal Pre-proof

Toll like receptors and their evolution in the lymnaeid freshwater snail species *Radix auricularia* and *Lymnaea stagnalis*, key intermediate hosts for zoonotic trematodes

Alexandra Juhász, Scott P. Lawton

PII: S0145-305X(21)00305-0

DOI: https://doi.org/10.1016/j.dci.2021.104297

Reference: DCI 104297

To appear in: *Developmental and Comparative Immunology*

Received Date: 7 April 2021

Revised Date: 11 October 2021

Accepted Date: 14 October 2021

Please cite this article as: Juhász, A., Lawton, S.P., Toll like receptors and their evolution in the lymnaeid freshwater snail species *Radix auricularia* and *Lymnaea stagnalis*, key intermediate hosts for zoonotic trematodes, *Developmental and Comparative Immunology* (2021), doi: https://doi.org/10.1016/j.dci.2021.104297.

Toll like receptors and their evolution in the lymnaeid freshwater snail species *Radix auricularia* and *Lymnaea stagnalis*, key intermediate hosts for zoonotic trematodes

Alexandra Juhász[1, 2], Scott P. Lawton[3]*

1. Institute of Medical Microbiology, Semmelweis University, H-1089 Budapest, Hungary

2. Department of Tropical Disease Biology, Liverpool School of Tropical Medicine, Liverpool, L3 5QA, UK

3. Epidemiology Research Unit (ERU) Department of Veterinary and Animal Sciences, Northern Faculty, Scotland's Rural College (SRUC), An Lòchran, 10 Inverness Campus, Inverness, IV2 5NA, UK

*****Correspondence: Scott P. Lawton, Epidemiology Research Unit (ERU) Department of Veterinary and Animal Sciences, Northern Faculty, Scotland's Rural College (SRUC), An Lòchran, 10 Inverness Campus, Inverness, IV2 5NA, UK

E-mail: scott.lawton@sruc.ac.uk

Abstract

One of the major evolutionarily conserved pathways in innate immunity of invertebrates is the toll-like receptor (TLR) pathway. However, little is known of the TLR protein family in gastropod molluscs despite their role in the transmission of human diseases, especially the common lymnaeid freshwater snail species *Radix auricularia* and *Lymnaea stagnalis*, key intermediate hosts of zoonotic trematodes. Using comparative genomics and gene prediction approaches utilising the freshwater snail *Biomphalaria glabrata* genome as a reference ten putative TLR proteins were identified in both *R. auricularia* and *L. stagnalis*. Phylogenetic analyses revealed that unlike other molluscs the lymnaeid species also possessed class 1 TLRs, previously thought to be unique to *B. glabrata*. Gene duplication events were also seen across the TLR classes in the lymnaeids with several of the genes appearing to exist as potential tandem elements in *R. auricularia*. Each predicted TLR was shown to possess the typical the leucine-rich repeat extracellular and TIR intracellular domains and both single cysteine clusters and multiple cysteine clusters TLRs were identified in both lymnaeid species. Principle component analyses of 3D models of the predicted TLRs showed that class 1 and 5 proteins did not cluster based on similarity of structure, suggested to be potential adaptation to a range of pathogens. This study provides the first detailed account of TLRs in lymnaeids and affords a platform for further research into the role of these proteins into susceptibility and compatibility of these snails with trematodes and their role in transmission.

Key words: Toll-like receptor, gastropod, *Radix auricularia*, *Lymnaea stagnalis*, gene duplication, leucine-rich repeat, phylogeny

Highlights:

- First attempt to predict and analyse putative TLR proteins in lymnaeid species
- Indentification of the occurance of molluscan class 1 TLRs in more species not only *Biomphalaria glabrata*
- Identification of both single cysteine clusters and multiple cysteine clusters TLRs in lymnaeids
- Gene duplication and structural radiation events of TLRs in lymnaeids could be indicative of adapation to a multiple pathogen environment

1. Introduction

As in all invertebrates, gastropods and other molluscs are dependent upon the genetically predetermined innate immune response, the first line of defence against pathogens and other environmental stresses (Schultz and Adema 2017). One of the major evolutionarily conserved pathways in innate immunity of invertebrates is the toll-like receptor (TLR) pathway, with the TLR family being one of the most intensively studied groups of proteins in immunology. The TLRs are trans-membrane proteins composed of an extracellular amino terminus and an amino-terminal leucine-rich repeat domain, responsible for pathogen recognition. Also, the TLRs have a conserved cytoplasmic carboxy-terminal intracellular Toll/interleukin-1 receptor (TIR) domain, which is responsible for signal transduction and activation of effector functions (Bowie and O'Neill 2000). These proteins are recognized as a major family of pattern recognition receptors, are homologues of the Toll protein that was first identified in *Drosophila* and discovered to play a role in pathogen responses (Hashimoto et al., 1988; Lemaitre et al., 1996). In molluscs the TLRs are expressed in haemocytes, the circulating phagocytic immune cells that are involved into the detection and destruction of parasites and pathogens (Brennan and Gilmore, 2018). However, comparatively fewer studies have been performed on the TLRs of molluscs relative to other invertebrates with most of the work detailing the expression and localisation of these receptors in economically important bivalve molluscs in particularly oysters (Brennan and Gilmore, 2018). Only recently have TLRs in gastropods, obligate intermediate hosts for trematodes, been explored in detail with the majority of work focused on the phylogenetically distinct heterobranch *Biomphalaria glabrata* and the prosobranchs *Oncomelania hupensis*, the snail intermediate hosts of agents of the neglected zoonotic tropical disease schistosomiasis, affecting in access of 200 million people in the tropics and sub tropics (Knight et al., 2014; Pila et al., 2016a,b,c; Zhao et al., 2018). The majority of the research undertaken in understanding the immune system of snails has focused on the interaction of the blood fluke *Schistosoma mansoni* and snail species within the genus *Biomphalaria* (Knight et al., 2014; Pila et al., 2016a,b,c; Adema et al., 2017; Zhao et al 2018). Within the *Biomphalaria-S. mansoni* system it has been shown that the success of fluke infection within an individual snail depends upon the efficacy of parasite-snail compatibility which is predetermined by parasite infectivity as well as snail defence capacity and specificity (Adema and Loker, 2015). These studies have illustrated that the TLRs, which are highly expressed in the snail's phagocytosing haemocyte cells, are key to haemocyte activation against pathogens and play an essential role in snail susceptibility to fluke infections (Adema et al., 2017; Pila et al., 2016c).

However, TLRs and components of the immune system of freshwater snails within the Lymnaeidae, which represent the one of the most important gastropod families owing to their medical and veterinary importance as intermediate hosts of zoonotic trematodes, remain largely understudied (Seppälä et al., 2021). Lymnaeid snails are the intermediate hosts to avian parasitic schistosomatid blood flukes the agent of human cercarial dermatitis (CD), caused by the penetration of the skin of cercariae released from the snail when in contact with the skin. Considered to be a re-emerging disease across Europe and an emerging disease globally infection presents with maculo-papulo-vesicular eruptions and eventually leading to itching, fever, swelling of the lymph nodes and even erythema and oedema (Lawton et al., 2014). This now also considerable experimental evidence that although many of the cercariae die in the skin several are able to survive to the schistosomula stage and migrate to lungs or can even become lost in the nervous system causing significant pathogenesis beyond those found in the skin (Horak et al., 2010). Also, lymnaeid snails are the intermediate host of several species of mammalian schistosomes across Asia, most notably *Schistosoma turkestanicum* a parasite also associated with CD in farm workers and fishing communities but also considered to be a major parasite of livestock particularly in sheep and cattle (Wang et al., 2009). In the Middle and Far East *S. turkestanicum* has been implicated in substantial loss in meat and milk yields and associated with significant morbidity and mortality in flocks and herds having considerable economic impact on the farming communities (Wang et al., 2009). In the past decade *S. turkestanicum* has also been identified in Central Eastern Europe and could represent a significant emerging disease of livestock (Juhász and Majoros, 2018). Similarly, lymnaeid snails represent a major group of vectors of food borne trematodiasis (FBT), including the zoonotic liver flukes *Fasciola gigantica* and *Fasciola hepatica* remain largely understudied. Fürst et al. (2020) estimated the global burden of FBT to approximate to 665 352 DALYs, a substantial socioeconomic issue to populations in endemic regions and a significant global health problem. Thus, understanding the molecular and genetic components of trematode-lymnaeid snail interactions not only provides deeper understanding of the molluscan immune system, but also provides a platform for the identification of markers which could provide insight into snail susceptibility in wild populations in endemic regions.

With the advent of extensive genome sequencing and detailed gene annotation the identification of immune genes and gene families has increased exponentially across many taxa particularly in the vertebrates, the nematodes and the arthropods. However, despite the

gastropods being considered the second most diverse taxa after the arthropods there remains relatively few reference gastropod genomes, with only the genome of *B. glabrata* having been extensively studied with a substantial number of putative genes annotated (Adema et al., 2017). Yet regardless of the importance of *R. auricularia* and *L. stagnalis* as disease vectors little is understood of the molecular and genetic components of their immune system and their interactions with pathogens. Whilst *Biomphalaria* falls within the family Planorbidae and *R. auricularia* and *L. stagnalis* fall within the family Lymnaeidae both closely related families fall within the Pulmonata group within the superorder Hygrophila. The close relationship between the Planorbidae and Lymnaeidae means that the genomic resources available for *B. glabrata* are the most appropriate for performing comparative and genome mining analyses of *R. auricularia* and *L. stagnalis.* By utilising the currently available genome assembly of *R. auricularia* and *L. stagnalis* the aim of this study was to identify putative members of the TLR gene family and to provide insights into the structure and composition of relationships of *R. auricularia* and *L. stagnalis* TLRs with those of *B. glabrata.*

2. Materials and Methods

2.1. Identification of scaffolds containing putative TLR genes from *R. auricularia* and *L. stagnalis*

In order to predict the TLR orthologs from the *R. auricularia* and *L. stagnalis* the *B. glabrata* genome was used as a reference, as it represents the most detailed and well annotated gastropod genome currently available and is also another Pulmonata aquatic snail, sharing several conserved protein families and pathways as with the lymnaids. Owing to the challenges in accuracy of gene annotation across fragmented genomes only the predicted protein sequences for complete *B. glabrata* TLRs (BgTLR) as specified in Adema et al. (2017) were retrieved from the current genome assembly available at VectorBase (https://vectorbase.org/vectorbase/app/record/organism/TMPTX_bglaBB02), which included BgTLR1 (BGLB022082), BgTLR7 (BGLB008408), BgTLR8 (BGLB010216), BgTLR9 (BGLB029612), BgTLR10 (BGLB028538), BgTLR11 (BGLB010218), BgTLR12 (BGLB038991), BgTLR13 (BGLB000615), BgTLR14 (BGLB035750), BgTLR15 (BGLB021504), BgTLR16 (BGLB039757), BgTLR19 (BGLB025483), BgTLR20 (BGLB023327), BgTLR21 (BGLB028788), BgTLR22 (BGLB020812), BgTLR23 (BGLB009964), BgTLR25 (BGLB005059), BgTLR26 (BGLB025331), BgTLR27 (BGLB021092), BgTLR28 (BGLB031420), BgTLR31 (BGLB040335), BgTLR34 (BGLB007854). Also, current reference proteins available on NCBI were for TLRs in *B. glabrata* were also collected to perform searches (Supplementary Table 1). The BgTLR reference protein sequences were used to identify scaffolds within the *R. auricularia* (GenBank acc: GCA_002072015) and the *L. stagnalis* (GenBank acc: GCA_900036025) genome which may contain putative TLR genes. The BgTLR were used to perform tBLASTn searches against the current genome assembles housed on NCBI (https://blast.ncbi.nlm.nih.gov/Blast). BLAST hits to scaffolds with the lowest e-values were considered to contain potential orthologs of TLRs and were used in further analyses to predict putative protein coding regions and subsequent protein sequences.

2.2. Prediction of putative TLR genes in *Radix auricularia* and *Lymnaea stagnalis*

Owing to the lack of detailed annotation of the *R. auricularia* and *L. stagnalis* genomes and a deficit in RNAseq reference data the *ab initio* gene prediction tool FGENESH, housed at on the Softberry computational biology web server (http://www.softberry.com/berry.phtml), was

used to predict putative TLR coding regions in the scaffolds previously identified by BLAST. The FGENESH is considered to be one of fastest and most accurate gene prediction tools and has been used previously to predict genes in other molluscs including the land air-breathing slug *Incilaria fruhstorferi* (Patnik et al., 2019) and the California sea hare *Aplysia californica* (Gorbushin et al., 2010). This approach also allowed predictions to be performed not only using the BgTLR protein sequences in an HMM protein-based gene prediction process but also by allowing customised gene prediction parameters to be enforced using gene prediction models for multiple invertebrate and vertebrate species. Overall gene predictions across the *R. auricularia* and *L. stagnalis* scaffolds were identical regardless of which gene prediction approach was employed. The resultant putative protein sequences were subjected to BLASTp searches on NCBI to provide an initial identification and to verify as a novel putative TLR protein sequence. To verify gene structure and accurately identify exon-intron structure for each of the predicted complete protein coding genes, putative complete TLR protein sequences were submitted to WebScipio (www.webscipio.org) (Hatje et al., 2011). This approach utilizes BLAT to perform initial protein-DNA spliced alignment searching for missing codons with preference given to those at splice sites and the adds nucleotides to the corresponding exons. This approach allows for the accurate identification of exon-intron boundaries within and across genome contigs (Hatje et al., 2011).

2.3. Phylogenetic analyses and classification of TLRs

In total 10 complete and unique TLRs were identified across both the *R. auricularia* and *L. stagnalis* scaffolds. In order to identify which class of TLRs the resultant proteins for *R. auricularia* (RaTLR) and *L. stagnalis* (LsTLR) belong to they were aligned against the BgTLR reference sequence for comparison to the same TLR classes as described by Adema et al. (2017). The TLR protein sequences form *B. glabrata*, *R. auricularia* and *L. stagnalis* were aligned using MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/) and resultant multiple sequence alignments were visualised and curated in BioEdit (Hall, 1999). Adema et al. (2017) described seven different classes of TLRs in *B. glabrata* components of which also appeared to be present in *A. californica.* To assess the occurrence of the TLR classes in *R. auricularia* and *L. stagnalis* a *p*-distance matrix was composed in MEGA X (Kumar et al., 2018) based on the TLR alignment as described above. MEGA X was also used to infer evolutionary relationships between the BgTLR, RaTLR and LsTLR using the Neighbour Joining (NJ) phylogenetic approach using the Poisson correction method, and 1000 bootstrap replicates

were used to provide nodal support across the tree. The resultant trees were then analysed using the Gene Duplication Wizard also implanted through MEGA X (Kumar et al., 2018; Zmasek and Eddy, 2001) to identify which branches in the phylogeny were the result of gene duplication events and those that had emerged through speciation. To provide a broader evolutionary perspective of RaTLRs and LsTLRs all predicated protein sequences representing TLR from all other gastropods were retrieved from NCBI, species included the freshwater snail species *O. hupensis*, *Pomacea canaliculate*, the marine snail species *Littorina littorea* and *Haliotis discus discus*, and the sea slugs *Elysia marginate*, *A. californica* and *Plakobranchus ocellatus*. Alignments and phylogenetic analyses were performed as descried previously using MEGA X but with out the gene duplication inference analyses.

2.4. TLRs structural elements analysis and identification

To ensure that the resultant putative RaTLRs and LsTLRs proteins were true representations of TLRs analyses of protein domain content and overall structure was performed to assess the occurrence of typical TLR characteristics. Initially, structural analyses employed by Zhao et al. (2018) were performed on each of the novel RaTLR and LsTLR protein sequences. Protein domains and motifs were predicted using the Simple Modular Architecture Research Tool (SMART) program (http://smart.embl-heidelberg.de/), which identified domains including signal peptides, the leucine-rich repeat (LLR) and TIR domains as well as identifying which parts of the protein were extracellular and intracellular and the region that passes through the cell membrane.

Toll like receptors are characterised according to the number of cysteine clusters which occur in their extracellular domain, with some TLR proteins containing single cysteine clusters (sccTLRs) and others containing multiple cysteine clusters (mccTLRs) having two or more cysteine clusters within the LLR extracellular domain. To identify and confirm the number of cysteine clusters within each of the RaTLRs and the LsTLRs the LRRfinder employed (http://www.lrrfinder.com/lrrfinder.php), (Offord et al., 2010). Finally, in order to determine the occurrence of the typical horse shoe shape in the RaTLRs and LsTLRs the three-dimensional (3D) structures of the proteins were created by the iterative implementation of multi-threading/assembly/refinement approach (Zhang, 2008) online server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/). The primary protein sequence of predicted TLRs were submitted into the I-TASSER web server and the best model with the

highest confidence score (C-score) was obtained. C-score was utilized to compute the quality of determined 3D protein models, where a higher value indicates a model with a high confidence (Zhang 2008). C-score is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations (Zhang 2008). Models were composed for each class of RaTLR and LsTLR as identified by the phylogenetic analyses. Structural differences of the resultant 3D models of the proteins were compared in DALI, the protein structure comparison server (Holm, 2020) and further PCA analyses was performed in PAST3 (Hammer et al., 2001) to assess if the putative TLRs also clustered based on structural similarity as well as sequence homology.

3. Results

3.1. Identification of scaffolds containing TLRs

Initial BLAST searches using both the BgTLRs from Adema et al (2017) and reference sequences derived from other studies identified 14 scaffolds from *R. auricularia* and a further 16 scaffolds from *L. stagnalis* genome assembles (Fig. 1, Supplementary table 1). Across both genomes most of the scaffolds were identified to share homology with a single TLR protein from the *B. glabrata* reference sequence. However, in *R. auricularia* scaffold MUZC01001691 had the top BLAST result for at least eight of the *B. glabrata* reference sequence, with scaffold FCFB01138811 also being the top BLAST result of eight reference TLRs in *L. stagnalis* (Fig. 1). Regression analyses of the association between scaffold length and BLAST hit frequency indicated there was no significant relationship between the two in either the *R. auricularia* or *L. stagnalis* assemblies. This illustrated that there was no bias to BLAST searches simply retrieving the longest sequences by default (Fig. 1). In the *R. auricularia* only seven scaffolds were shown to produce complete putative TLRs based on the FGENESH gene prediction, with scaffold MUZC01004652 containing three and MUZC01000569 containing a further two complete TLRs. This was not the case for *L. stagnalis* with FGENESH being able to predict TLRs from 10 of the identified scaffolds. From each of those *L. stagnalis* scaffolds only a single complete TLR could be predicted (Supplementary table 1). In both *R. auricularia* and *L. stagnalis* there was no consistent pattern in gene structure for the predicted TLR protein coding genes other than half of the total predicted complete proteins corresponded to large single exon genes that ranged from 2424bp-3387bp (Supplementary Fig. 2 and 3). For those genes that did have exons there was no obvious pattern of intron-exon number or size. Some

genes contained a single intron with five introns being the maximum predicted. Interestingly, in *R. auricularia* in the TLRs predicted on MUZC01004652 are suspected to be duplicated two of the genes had almost identical length but the third contained two introns (Supplementary Fig. 2).

3.2 Classification and phylogeny of the toll like receptors in *R. auricularia* and *L. stagnalis*

To identify which class each of the complete RaTLRs and LsTLRs belong to phylogenetic analyses was performed using the BgTLRs from Adema et al. (2017) as they had already been categorised from class 1 – 7 (Fig. 2). The phylogeny identified five distinct clusters with RaTLRs and LsTLRs clustering within class 1, 4, 5, 6 and 7. Neither *R. auricularia* or *L. stagnalis* had any TLRs represented in class 2 or 3 relative to those in *B. glabrata*. However, interestingly the *B. glabrata* TLRs from classes 1, 2 and 3 appeared to be exceptionally closely related with classes 1 and 2 being phylogenetically bracketed by members of class 1 and BgTL20 and BgTLR28 appearing basal to classes 1, 2 and 3 despites being originally classified as class 1 in Adema et al. (2017) (Fig. 2). Across the tree both RaTLRs and LsTLRs clustered together within lymnaeid specific subclades and in class 1 the RaTLRs and LsTLRs formed two paraphyletic subclades to the *B. glabrata,* with one of the subclades appearing basal to the rest of the clade. One of the subclades sat firmly within class 1 and indicated a gene duplication event in *R. auricularia* with three putative TLRs from scaffold MUZC01004652 clustering together (Fig. 2). Interestingly, in class 5 there did not appear to be specific lymnaeid clades forming with the majority of RaTLRs and LsTLRs being scattered amongst the TLRs from *B. glabrata* (Fig. 2). Class 7 indicated the occurrence of a gene duplication event in *L. stagnalis* with the only known predicted TLR from *L. stagnalis* (Sappala et al., 2021) forming a paraphyletic lineage to the main class 7 clade which contained TLRs form *L. stagnalis*, *R. auricularia* and *B. glabrata* (Fig. 2).

Phylogenetic analyses across the gastropods indicated that the 7 classes identified in *B. glabrata* were in fact shared across the gastropods with again only *R. auricluaria*, *L. stagnalis* and *B. glabrata* having representatives in class 1, and classes 2 and 3 being specific to *B. glabrata* (Fig. 3). The clade corresponding to class 4 contained TLR homologs form *L. stagnalis*, *R. auricularia* and *B. glabrata* but also the sea slugs *P. ocellatus* and *E. marginate*. Class 5 appeared to contain TLRs from all the freshwater snails analysed including *L. stagnalis*, *R. auricularia, B. glabrata, O. hupensis* and *P. canaliculate* as well as sea slugs *A. californica*

and *E. marginate*. The clade representing class 6 was the largest clade containing 32 TLRs from six snail species. This clade contained only TLRs from marine and freshwater snails with 12 from *P. canaliculate*, 2 from *L. stagnalis*, 6 from *O. hurpensis*, 2 from *H. discus discus*, 9 from *L. littorea* and a single TLR from *B. glabrata*. Class 7 appeared to be composed of TLRs from the sea slug species *E. marginate*, *A. californica* and *P. ocellatus* but did contain a specific pulomonata snail linage which represented TLRs from *L. stagnalis*, *R. auricularia* and *B. glabrata* (Fig. 3).

3.3 Structural comparisons of the TLR in *R. auricularia* and *L. stagnalis*

The SMART analyses of the putative RaTLR and LsTLRs predicted a typical transmembrane TLR structure. Each protein was shown to have an extracellular domain containing numerous LRR motifs in the N-terminal region and a single endocellular conserved TIR domain in the C terminus. The RaTLR and LsTLR sequence length ranged from 727 to 1119 amino acids (aa) and typically signal peptides were identified in the N-terminal regions of both *R. auricularia* and *L. stagnalis* (Fig. 4). However, several TLRs were shown to be missing a signal peptide with four representatives from *L. stagnalis*, two of which been shown from class 1, 1 from class 5 and 1 from class 6, whereas *R. auricularia* showed only 2 TLRs to lack a signal peptide from class 1 and class 4. The I-TASSER 3D predictions of RaTLRs and LsTLRs showed that all of the putative proteins had the typical horseshoe shape as seen in other taxa with the large curved extracellular LRR region and the endocellular TIR domain (Fig. 5, Supplementary fig. 3). Principle component analyses based on comparisons of the 3D structures of the TLRs did indeed show some evidence of clustering the proteins to their correct classes based on structure (Fig. 6). Although specific clusters for Class 4, 5, 6 and 7 could be identified based on shape, TLRs from class 1 did appear to be scattered amongst the classes indicated substantial variation in structure. All members of the RaTLR and LsTLR were shown to contain a variable number of LRR domains ranging from 2 to 9 repeats within the extracellular region. There did not appear to be any uniformity in the position or in the number of LRRs within each of the TLR classes (Fig. 4). However there did appear to be some similarities LRR content between the TLRs predicted from *R. auricularia* or *L. stagnalis* that represented orthologs as seen in class 1 and 5 but especially in class 7. A total of eight sccTLRs and two mccTLRs were predicted in both *R. auricularia* and *L. stagnalis*, with mccTLRs found in both species in class 4 and 7 as identified by the phylogenetic analyses (Supplementary table 2). In *L. stagnalis* the mccTLRs sat in distinct positions from 549-701aa and then from 824-1129aa, unlike those found in class

7 which sat next to each other positions 385-568aa and 569-786aa within the peptide sequence. This pattern was also seen in *R. auricularia* with the distinct separation of the cystine clusters at 342-494aa and 617-922aa in mccTLRs in class 4 but sat next to each other in class 7 at positions 361-568aa and 569-786 along the peptide sequence (Supplementary table 2 and Supplementary Fig. 3).

4. Discussion

In recent years owing to the ever-increasing availability of high-quality reference genomes for non-model organisms the identification of genes and the exploration of gene families has become more efficient and a major area of research in invertebrate biology. In this current study ten putative complete TLR genes were identified for the first time in the lymnaeid snails *R. auricularia* and *L. stagnalis* based on gene prediction models designed to predict genes based on homology searches, a standard approach for organisms ranging from animals and plants which lack detailed transcriptomic data (Adema et al., 2017; Deng et al., 2014; Patnaik et al., 2019; Yu et al., 2020; Terada et al., 2016). Predicted TLR protein sequences that were identified in this study were directly comparable with those identified in *B. glabrata* and were homologous to those within TLR classes 1, 4, 5, 6 and 7 as described in Adema et al. (2017). This provided the validity of the accuracy of the gene prediction method in this current study but the lack of TLRs within classes 2 and 3. This may be a true indication of the lack of such TLRs in lymnaeids as the phylogenetic analyses showed these TLRs to be absent across all the gastropod species analysed in this study, however, it is important to note that this may also be the result of a paucity in the completeness of *R. auricularia* and *L. stagnalis* reference genomes, considered to be an issue of reference genomes for non-model organisms (Deutekom et al., 2019). The phylogenetic analyses showed the RaTLRs and the LsTLRs to consistently form lymnaeid specific subclades within classes TLR 1, 4, 5, 6, and 7 when compared with those from *B. glabrata*. Historically, *B. glabrata* has been shown to be the only mollusc possesses TLRs in classes 1-3, suggested to be the consequence of recent duplication events with species specific roles (Adema et al., 2017). However, the occurrence of two TLRs from *L. stagnalis* and four from *R. auricularia* clustering within the *B. glabrata* class 1 clade raises interesting questions about their occurrence across the snail superorder Hygrophila which contain both Planorbidae and Lymnaeidae which are closely related are form sister lineages within the superfamily Lymnaeoidea (Saadi et al., 2020). Adema et al. (2017) noted the broad radiation of class 1 TLRs in B. glabrata and suggested that this may indeed have been in response to the

complex relationship with *Schistosoma* blood flukes. The occurrence of class 1 TLRs in *L. stagnalis* and *R. auricularia* may also indicative that these TLRs are in fact specific to the Lymnaeoidea and have evolved owing to their complex interactions not only with schistosomatids but other Strigeatida parasites too. All the snail species within the Lymnaeoidea play host to different Strigratida parasites, which include the avian and mammal schistosomes as well the economically important Diplostomida and only through continued comparative genome analyses as genomes for more snail species become available will it be possible to assess the radiation of class 1 TLRs in relation to parasite diversity within the Lymnaeoidea and across the Hygrophila.

Unsurprisingly, the phylogenetic analyses also indicated that speciation played a key role in the diversification of the TLRs between *R. auricularia* and *L. stagnalis* and the precursors of the genes would have been present in an ancient common ancestor (Innan and Kondrashov, 2010; Nei and Rooney, 2005). However, both *R. auricularia* and *L. stagnalis* did appear to show class specific gene duplication events. *Radix auricularia* showed to have TLR duplication events in class 1 and 5 which was also supported with the multiple BLAST hits of the same *B. glabrata* homologs to different parts of the same scaffold. Similarly, *L. stagnalis* also appeared to have duplication events in class 1, 4, 5 and 6 but from multiple scaffolds. Gene duplication and expansion events in immune genes are considered to be important sources of evolutionary novelty as natural selection maintains duplicated genes as a consequence of functional divergence (Innan and Kondrashov, 2010; Nei and Rooney, 2005). In other invertebrate and vertebrate species TLRs have been shown to have undergone expansive gene duplication events and that such genes often exist in tandem clusters. This has recently been shown in some species of bivalve molluscs such as *Crassostrea gigas* (Zhang et al., 2015) and *Ruditapes decussatus* (Batista et al., 2019) with TLRs not only duplicating in tandem repeat but each TLR within a specific lineage having a unique function involved in recognising specific pathogens, involved in detoxification and some indication of specific roles in development. The TLRs found on the same scaffolds in *R. auricularia* do in fact occur closely together and would represent a tandem duplication event and clustered closely in the phylogenetic analyses. The fresh water snails *B. glabrata* and *O. hupensis* have also been shown to have under gone major expansion events of their TLR repertoire, hypothesised to be linked to the wide range of pathogens found in dynamic freshwater ecosystem and is likely to be the same selective force driving the diversification in the lymnaeids (Adema et al., 2017; Knight et al., 2014; Pila et al.,

2016a,b,c; Zhao et al 2018). As discussed by Zhang et al. (2015), as in oysters, it is suspected that the total numbers of TLRs and those in tandem linkage repeats in lymnaeids is massively underestimated by this study probably as result of scaffold length and assembly quality, but also as a result of the loss of minimally diverged paralogs during genome assembly (Roach et al., 2005). This may account for the lack of identification of tandem linkages in *L. stagnalis* but is further indication of the requirement for a detailed account of the evolution of TLR genes and proteins across the gastropods supported by detailed experimental work. The phylogenetic analyses across gastropods also showed there to be extensive radiation in the TLR classes between species specifically within in class 6 which represented each snail species and family. Studies of radiation in TLR classes in teleost fish (Solbakken et al., 2017) and cetaceans (Shen et al., 2012) have linked the divergence to the utilisation of different environments and habitats and the different pathogens that may be encountered. These same selective processes may account for the radiation in the snails for class 6 TLRs with *P. canaliculate*, *L. stagnalis*, *B. glabrata* predominantly freshwater, *O. hurpensis* being associated with freshwater mud banks, with *H. discus discus* and *Li. littorea* being associated with coastal marine environments. Similarly, as described by Solbakken et al. (2017) and Shen et al. (2012) geographical distribution between habitats will have an implication in radiation for example *P. canalculate* is found across South America and *L. stagnalis* across Northern Europe and Asia, so although both occupying the same type of environment subtle differences associated with local pressures could also impact on the diversity of TLRs. Interestingly, Seppälä et al. (2021) illustrated differentiation in transcription of other immune genes in *L. stagnalis* from different populations and environments illustrating that immune response to pathogens was associated with environmental conditions differed between populations in different habitats. Also, the study showed that although the components of the immune system were homologous between species, *B. glabrata* and *Li. littorea*, there were specific differences again considered to be associated with the different environments to which these species were inhabiting. These same factors could be driving the genetic diversity of class 6 TLRs not only across the gastropods but also within species, however detailed evolutionary and population genetic analyses is required to address this fully.

The TLRs predicted from *R. auricularia* and *L. stagnalis* all showed the typical characteristics of TLRs with an extensive LRR dense ectodomain, a short transmembrane region and the cytoplasmic TIR region. The 3D modelling of each of these proteins also revealed that they

had the typical horseshoe shape characteristic of the other known TLRs with some subtle differences in shape of the ectodomain. For both *R. auricularia* and *L. stagnalis* there appeared to be high variation in the distribution of LRR regions between paralogs within each of the classes, particularly class 1 and 5. Similarly, PCA analyses based on the 3D structure of the TLRs could not resolve classes 1 or 5 into single class specific clusters further indicating a considerable variation in shape and structure of the TLRs within the classes. The TLRs within class 1 and 5 had high levels of variation in the content and distribution of the LRR regions with little consistency between paralogs. This is a pattern also described in the *B. glabrata* class 1 and the variability between paralogs has been suggested to be associated with ligand binding to highly variable pathogen molecules or TLR dimerization as described in other organisms particularly mammals (Adema et al., 2017; Wlasiuk and Nachman, 2010).

Historically, TLRs have been identified into two prototypical types, those with a single cysteine cluster (sccTLR) and those with multiple cysteine clusters (mccTLR). The sccTLR are common throughout all animal life but the mccTLRs appear to be relegated to the protosomes including both the Ecdysozoa and the Lophotrochozoa (Brennan and Gilmore, 2018). In molluscs major gene expansion events have been predominantly seen in both TLR types with the *B. glabrata* snail estimated to encode 25 sccTLRs and two mccTLRs, and the oyster *C. gigas* estimated to encode four mccTLRs, five sccTLRs and a further 74 variants of sccTLRs and mccTLRs (Adema et al., 2017; Zhang et al., 2015). As in *B. glabrata* both *R. auricularia* and *L. stagnalis* both had an abundance of sccTLRs which had appeared to undergo major duplication events as has been described previously. However, also similar to *B. glabrata*, the lymnaeid snails also coded for two mccTLRs found within classes 4 and 7, further supporting the concept of Lymnaeidae specific patterns of TLR divergence causing them to be distinct from those seen bivalves as highlighted by Brennan and Gilmore (2018). It is also important to note that the mccTLR expression has been illustrated to be specifically associated with trematode infection (Pila et al., 2016; 2017). Gene knockdown studies of mccTLR genes in schistosome resistant strains of *B. glabrata* have shown to render the snails susceptible to *Schistosoma mansoni* (Pila et al., 2016; 2017). To date there have been no detailed studies on the relationships between zoonotic trematode species, their snail intermediate hosts and the snail TLR genes that could play a major role in the epidemiology of such parasites.

Molluscan TLRs are primarily expressed in the haemocyte immune cells and are involved in

pathogen detection and destruction and crucial in determinant susceptibility of snails to fluke infection and ultimately transmission (Brennan and Gilmore, 2018). Both *R. auricularia* and *L. stagnalis* have been shown to be crucial in the transmission of food borne trematodes and with continued detailed genomic analyses will a detailed account of the genetic basis of the parasite-host interface be revealed. Novel genome lead antiparasitic control methods are beginning to be devised based on the development of genetically engineered snails to aid in the control of schistosomiasis, with the mind the technology could be applied to the control of parasites for both human and veterinary health (Maier et al., 2019). Although the authors of this study agree that further validation of the results of TLRs in lymnaeid snails' experimental approaches are required. However, the approaches used for gene/protein prediction here provide a robust platform for initial identification of crucial immune genes and comparative immunogenetic studies in molluscs and other neglected groups

Acknowledgments

This research was supported by a grant MÁEÖ2018-2019/279479, EFOP-3.6.3-UK-2020, NTP-NFTÖ-19-B-0165, NTP-NFTÖ-20-B-0099 The authors thank the help by Fish Pathology and Parasitology Research Team in Institute for Veterinary Medical Research
Centre for Agricultural Research in collecting data in I-TASSER.

References

Adema, C.M., Hillier, L.W., Jones, C.S., Loker, E.S., Knight, M., et al., 2017. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. Nat. Commun. 16, 15451. https://doi: 10.1038/ncomms15451.

Adema, C.M., Loker, E.S., 2015. Digenean-gastropod host associations inform on aspects of specific immunity in snails. Dev. Comp. Immunol. 48, 275-83. https://doi: 10.1016/j.dci.2014.06.014.

Batista, F.M., Churcher, A.M., Manchado, M., Leitão, A., Power, D.M., 2019. Uncovering the immunological repertoire of the carpet shell clam *Ruditapes decussatus* through a

transcriptomic-based approach. Aquac. Fish. 4, 37-42. https://doi.org/10.1016/j.aaf.2018.03.003.

Brennan, J.J., Gilmore, T.D., 2018. Evolutionary Origins of Toll-like Receptor Signaling. Mol Biol. Evol. 35, 1576-1587. doi: 10.1093/molbev/msy050.

Bowie, A., O'Neill, L.A., 2000. The interleukin-1 receptor/Toll-like receptor superfamily: signal generators for pro-inflammatory interleukins and microbial products. J. Leukoc. Biol. 67, 508–514.

Correa, A., Escobar, J.S., Noya, O., Velasquez, L.E., Gonzalez-Ramirez, C., Hurtrez-Boussès, S. and Pointier, J.-P., 2011. Morphological and molecular characterization of Neotropic Lymnaeidae (Gastropoda: Lymnaeoidea), vectors of fascioliasis. Infect. Genet. Evol. 11, 1978–1988.

Deng, Y., Lei, Q., Tian, Q., Xie, S., Du, X., Li, J., Wang L., Xiong, Y., 2014. *De novo* assembly, gene annotation, and simple sequence repeat marker development using Illumina paired-end transcriptome sequences in the pearl oyster *Pinctada maxima*. Biosci. Biotechnol. Biochem. 78, 1685-1692, https://doi: 10.1080/09168451.2014.936351

Deutekom, E. S., Vosseberg, J., van Dam, T., Snel, B., 2019. Measuring the impact of gene prediction on gene loss estimates in Eukaryotes by quantifying falsely inferred absences. PLoS Comput. Biol. 15, e1007301. https://doi.org/10.1371/journal.pcbi.1007301

Fürst, T., Keiser, J., Utzinger, J., 2012. Global burden of human food-borne trematodiasis: a systematic review and meta-analysis. Lancet. Infect. Dis. 12, 210-21. https://doi: 10.1016/S1473-3099

536 Gorbushin, A.M., Panchin, Y.V., Iakovleva, N.V., 2010. In search of the origin of FREPs:
537 characterization of *Aplysia californica* fibrinogen-related proteins. Dev. Comp. Immunol. 34,
538 465-73. https://doi: 10.1016/j.dci.2009.12.007

539 Hatje, K., Keller, O., Hammesfahr, B., Pillmann, H., Waack, S., Kollmar., 2011. Cross-
540 species protein sequence and gene structure prediction with fine-tuned Webscipio 2.0 and
541 Scipio. BMC Res. Notes 4, 265 https://doi.org/10.1186/1756-0500-4-265
542
543 Hammer, Ø., Harper, D.A.T., and Ryan P.D., 2001. PAST: Paleontological Statistics
544 Software Package for Education and Data Analysis. Palaeontol. Electron. 4, 9.
545
546 Hashimoto, C., Hudson, K.L. and Anderson, K.V., 1988. The Toll gene of *Drosophila*,
547 required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein.
548 Cell. 52, 269-79

549

550 Holm, L., 2020. DALI and the persistence of protein shape. Protein Sci. 29, 128-140.

551

552 Horák, P., Kolářová, L., 2011. Snails, waterfowl and cercarial dermatitis. Freshwater Biol. 56,
553 779-790. 10.1111/j.1365-2427.2010.02545.x
554
555 Innan, H., Kondrashov, F., 2010. The evolution of gene duplications: classifying and
556 distinguishing between models. Nat. Rev. Genet. 11, 97–108.
557
558 Juhász, A., Majoros, G., 2018. Investigations on the distribution of *Schistosoma turkestanicum*
559 Skrjabin, 1913 (Trematoda: Schistosomatidae) infection of red deer in Hungary and a
560 combined method for the detection of *S. turkestanicum* eggs in droppings. Acta Vet. Hung. *66*,
561 587–606. https:// doi: 10.1556/004.2018.052

562

563 Knight, M., Arican-Goktas, H.D., Ittiprasert, W., Odoemelam, E.C., Miller, A.N., Bridger,
564 J.M., 2014. Schistosomes and snails: a molecular encounter. Front. Genet. 5, 230.
565

566 Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular
567 Evolutionary Genetics Analysis across computing platforms. Mol. Biol. Evol. 35**,**1547-1549.

568

569 Lawton, S. P., Lim, R. M., Dukes, J. P., Cook, R. T., Walker, A. J., Kirk, R. S., 2014.
570 Identification of a major causative agent of human cercarial dermatitis, *Trichobilharzia franki*
571 (Müller and Kimmig 1994), in southern England and its evolutionary relationships with other
572 European populations. Parasit. Vectors. 7, 277. https://doi.org/10.1186/1756-3305-7-277

573

574 Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J.-M., Hoffman, J.A., 1996. The dorsoventral
575 regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in
576 *Drosophila* adults. Cell. 86, 973-83.

577

578 Mas-Coma S, Valero MA, Bargues MD. 2009. *Fasciola*, lymnaeids and human fascioliasis,
579 with a global overview on disease transmission, epidemiology, evolutionary genetics,
580 molecular epidemiology and control. Adv. Parasitol. 69, 41–46.

581

582 Nei, M., Rooney, A.P., 2005. Concerted and birth-and-death evolution of multigene families.
583 Annu. Rev. Genet. 39, 121.

584

585 Offord, V., Coffey, T.J., Werling, D., 2010. LRRfinder: a web application for the identification
586 of leucine-rich repeats and an integrative Toll-like receptor database. Dev. Comp. Immunol.
587 34, 1035-1041.

588

589 Patnaik, B.B., Chung, J.M., Hwang, H.J., Sang, M.K., Park, J.E., Min, H.R., Cho, H.C.,
590 Dewangan, N., Baliarsingh, S., Kang, S.W., Park, S.Y., Jo, Y.H., Park, H.S., Kim, W.J., Han,
591 Y.S., Lee, J.S., Lee, Y.S., 2019. Transcriptome analysis of air-breathing land slug, *Incilaria*
592 *fruhstorferi* reveals functional insights into growth, immunity, and reproduction. BMC
593 Genomics. 26, 154. https://doi: 10.1186/s12864-019-5526-3.

594

595 Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating
596 signal peptides from transmembrane regions. Nat Methods. 8, 785–786. https://doi:
597 10.1038/nmeth.1701.
598
599 Pila, E.A., Gordy, M.A., Phillips, V.K., Kabore, A.L., Rudko, S.P., Hanington, P.C., 2016a.
600 Endogenous growth factor stimulation of haemocyte proliferation induces resistance to
601 *Schistosoma mansoni* challenge in the snail host. Proc. Natl. Acad. Sci. U. S. A. 113, 5305–
602 5310.
603
604 Pila, E.A., Sullivan, J.T., Wu, X.Z., Fang, J., Rudko, S.P., Gordy, M.A., Hanington, P.C.,
605 2016b. Haematopoiesis in molluscs: a review of haemocyte development and function in
606 gastropods, cephalopods and bivalves. Dev. Comp. Immunol. 58, 119–128.
607
608 Pila, E.A., Tarrabain, M., Kabore, A.L., Hanington, P.C., 2016c. A novel toll-like receptor
609 (TLR) influences compatibility between the gastropod *Biomphalaria glabrata*, and the
610 digenean trematode *Schistosoma mansoni*. PLoS Pathog. 12, e1005513.
611
612 Roach, J.C., Glusman, G., Rowen, L., Kaur, A., Purcell, M.K., Smith, K.D., Hood, L.E.,
613 Aderem, A., 2005. The evolution of vertebrate Toll-like receptors. Proc Natl Acad Sci U S A.
614 102, 9577-82. https://doi: 10.1073/pnas.0502272102.
615
616 Saadi, A.J., Davison, A., Wade, C.M., 2020. Molecular phylogeny of freshwater snails and
617 limpets (Panpulmonata: Hygrophila). Zool. J. Linn. Soc. 190, 518-531.
618 https://doi.org/10.1093/zoolinnean/zlz177
619
620 Seppälä, O., Walser, J. C., Cereghetti, T., Seppälä, K., Salo, T., Adema, C. M., 2021.
621 Transcriptome profiling of *Lymnaea stagnalis* (Gastropoda) for ecoimmunological
622 research. BMC Genomics. *22*, 144. https://doi.org/10.1186/s12864-021-07428-1
623
624 Schultz, J.H., Adema, C.M., 2017. Comparative immunogenomics of molluscs. Dev. Comp.
625 Immunol. 75, 3–15.

Shen, T., Xu, S., Wang, X., Yu, W., Zhou, K., & Yang, G., 2012. Adaptive evolution and functional constraint at TLR4 during the secondary aquatic adaptation and diversification of cetaceans. BMC Evol. Biol. 12, 39. https://doi.org/10.1186/1471-2148-12-39

Solbakken, M.H., Voje, K.L., Jakobsen, K.S., Jentoft, S., 2017. Linking species habitat and past palaeoclimatic events to evolution of the teleost innate immune system. Proc. Biol. Sci. 284, 20162810. https://doi.org/10.1098/rspb.2016.2810

Terada, D., Kawai, F., Noguchi, H., Unzai, S., Hasan, I., Fujii, Y., Park, S-Y., Ozeki, Y., Tame, J.R.H., 2016. Crystal structure of MytiLec, a galactose-binding lectin from the mussel *Mytilus galloprovincialis* with cytotoxicity against certain cancer cell types. Sci Rep. 6:28344.

Wang, C. R., Chen, J., Zhao, J. P., Chen, A. H., Zhai, Y. Q., Li, L., Zhu, X. Q., 2009. *Orientobilharzia* species: neglected parasitic zoonotic agents. Acta trop. 109, 171–175. https://doi.org/10.1016/j.actatropica.2008.11.009

Wlasiuk, G., Nachman, M.W., 2010. Adaptation and constraint at Toll-like receptors in primates. Mol. Biol. Evol. 27, 2172-2186.

Yu, X., Qin, Q., Wu, X., Li, D., Yang, S., 2020. Genetic localization of the SPC gene controlling pod coiling direction in *Medicago truncatula*. Genes Genomics. 42, 735-742. https://doi: 10.1007/s13258-020-00947-3.

Zhang, Y. 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 9, 40

Zhang, Y., He, X., Yu, F., Xiang, Z., Li, J., Thorpe, K.L., Yu, Z., 2015. Characteristic and functional analysis of toll-like receptors (TLRs) in the lophotrocozoan, *Crassostrea gigas*, reveals ancient origin of TLR-mediated innate immunity. PLoS One. 8, e76464.

Zhao, Q.P., Gao, Q., Zhang, Y., Li, Y.W., Huang, W.L., Tang, C.L., Dong, H.F., 2018. Identification of Toll-like receptor family members in *Oncomelania hupensis* and their role in defence against *Schistosoma japonicum*. Acta Tropica. 181, 69-78.

Zmasek, C., Eddy, S., 2001. A simple algorithm to infer gene duplication and speciation events on a gene tree. Bioinformatics. 17, 821-828.

Figures:

Figure 1: Summary of scaffolds identified through BLAST as containing putative TLRs for *Radix auricularia* and *Lymnaea stagnalis*. Where Ai) is a bar chart illustrating which scaffolds were identified through BLAST against the *R. auricularia* genome assembly, the frequency by which the scaffold was hit by a different BgTLR reference sequence, and where it was possible to predict complete putative TLRs. Aii) is a regression analyses that illustrates the lack of relationship between scaffold length and frequency of BLAST hits. B i and ii represent the same as described above but for *L. stagnalis*.

Figure 2: Phylogenetic reconstruction of TLRs identified in *Radix auricularia* and *Lymnaea stagnalis* in relationship to those characterised in *Biomphalaria glabrata,* indicating both speciation and gene duplication events. A distance-based neighbour joining phylogenetic analyses indicating the identification and evolutionary position of TLRs from *R. auricularia* and *L. stagnalis* within the classes specified for *B. glabrata.* Nodal bootstrap supports are shown at >50 across the phylogeny. Blue lined diamonds represent gene duplication events with red lined diamonds representing speciation events.

Figure 3: Phylogenetic reconstruction of TLRs from across the Gastropoda illustrating the diversity and variation within each class of TLRs. A distance-based neighbour joining phylogenetic analyses indicating the radiation of the TLR classes identified in *Biomphalaria glabrata* and their relationships across the gastropods. Nodal bootstrap supports are shown at >50 across the phylogeny

Figure 4: Linear schematics of TLR architecture identified in *Radix auricularia* and *Lymnaea stagnalis*. Where dark blue rectangles represent trans-membrane regions while green bands indicate LRRs involved in pathogen recognition. Light grey and blue circles represent N terminal and C terminal LRRs respectively, thin orange bars represent signal peptides and thin gray bars segments of low compositional complexity. TLRs consist of an intracellular TIR domain (blue hexagon)

Figure 5: Three dimensional models of the structure of exemplar TLRs from *Radix auricularia* and *Lymnaea stagnalis*. Where A) represents a class 1 TLR from *L. stagnalis* and B) represents a class 1 TLR from *R. auricularia.* Both indicate a typical horse-shoe shape as seen in other taxa. Also, highlight a single cysteine cluster are highlighted in yellow on the protein structure.

Figure 6: Principle component analyses based on comparisons of 3D structure of TLRs predicted from *Radix auricularia* and *Lymnaea stagnalis* genomes. The PCA analyses is based on comparisons performed in DALI which quantified structural differences calculated from direct comparisons of each 3D model from across all the TLR classes.

Supplementary data

Supplementary Table 1: BLAST results and FGENESH gene position prediction of toll like receptor genes in the genomic scaffolds of the lymnaeid snails *Radix auricularia* and *Lymnaea stagnalis*

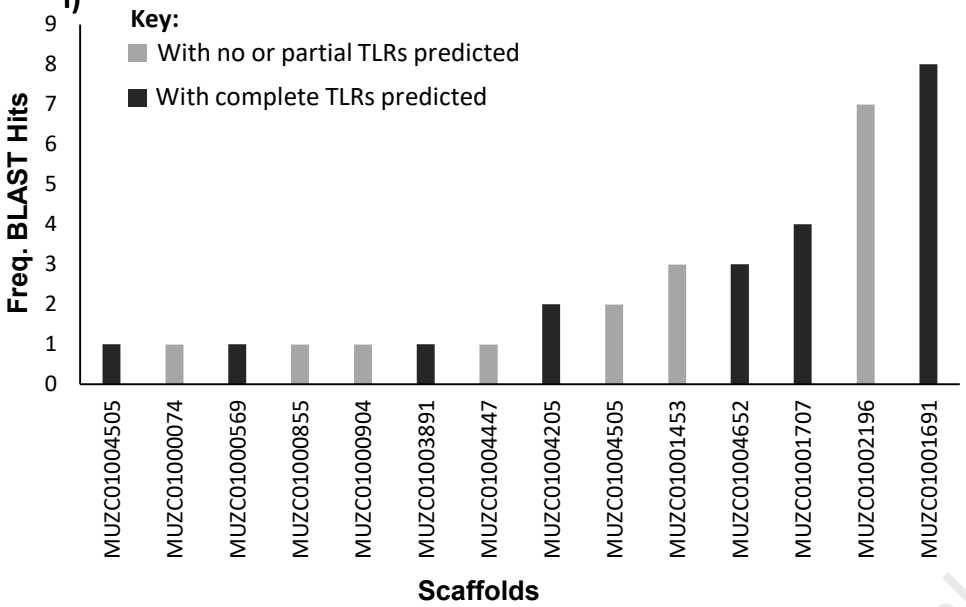
Supplementary Table 2: Identification of LRR regions in predicted toll like receptor protein sequences for the *Lymnaea stagnalis* and *Radix auricularia*

Supplementary figure 1: Schematic of predicted structure of TLR genes in *Radix auricularia*. Where dark grey rectangles represent coding regions (exons) and light grey represent representing intronic regions (introns). The image is not to scale in order to emphasise

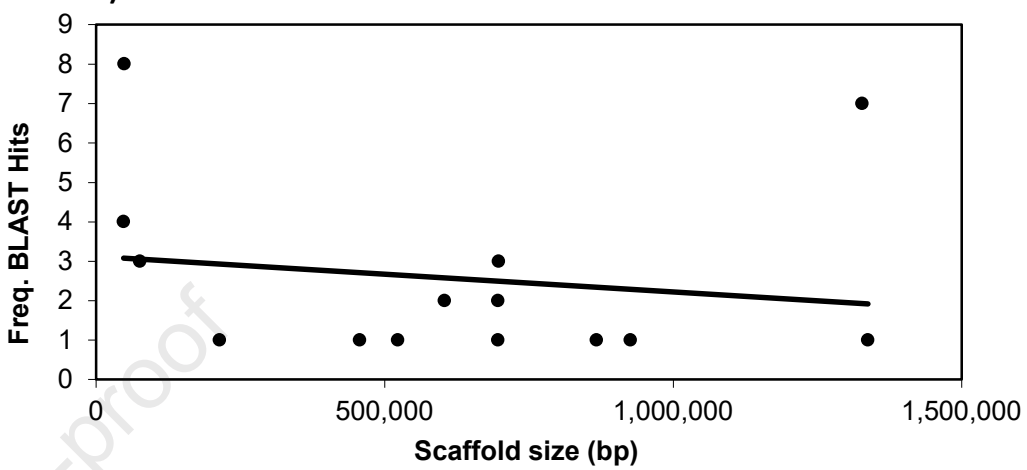
727 coding regions
728
729 **Supplementary figure 2: Schematic of predicted structure of TLR genes in *Lymnaea***
730 ***stagnalis***. Where dark grey rectangles represent coding regions (exons) and light grey represent
731 representing intronic regions (introns). The image is not to scale in order to emphasise coding
732 regions
733
734 **Supplementary figure 3: Predicted 3D structures of representations of putative *Radix***
735 ***auricularia* and *Lymnaea stagnalis* TLRs from each of the TLR classes.** Comparisons
736 between *R. auricularia* and *L. stagnalis* within each class indicate a typical horse-shoe shape
737 as seen in other taxa. Also, those classes which are represented by single and multiple cysteine
738 clusters are highlighted in yellow. Note the mccTLRs of class 7 are sat next to each other so
739 appear as one continuous coloured section.
740

A) *Radix auricularia*

i)

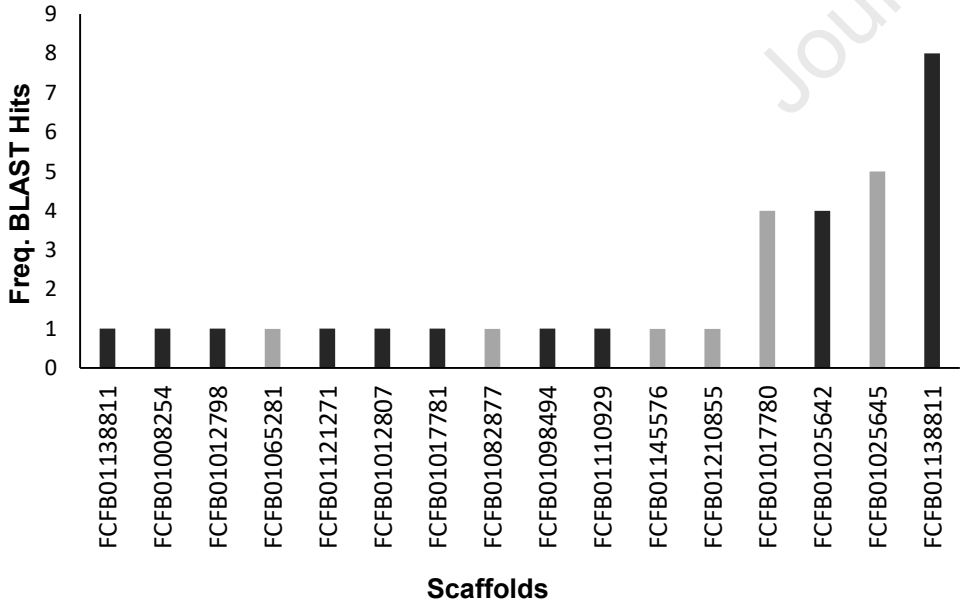


ii)

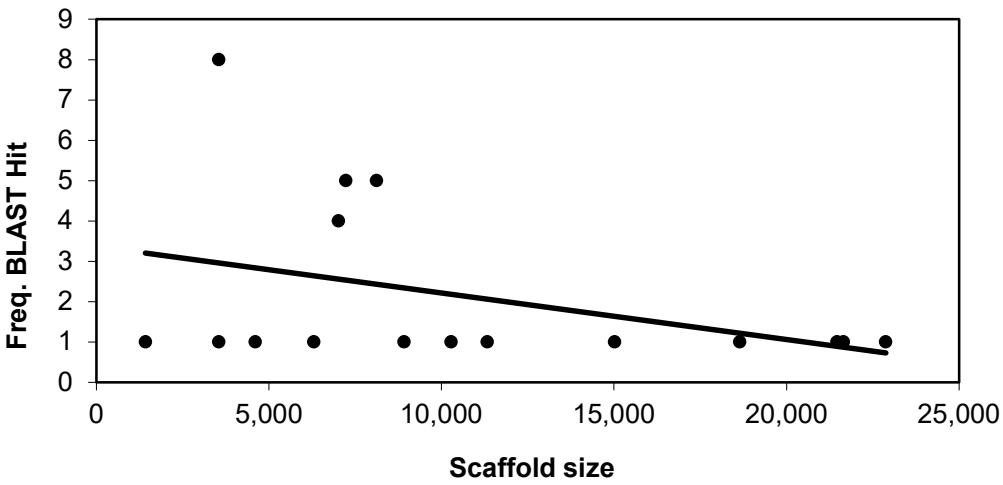


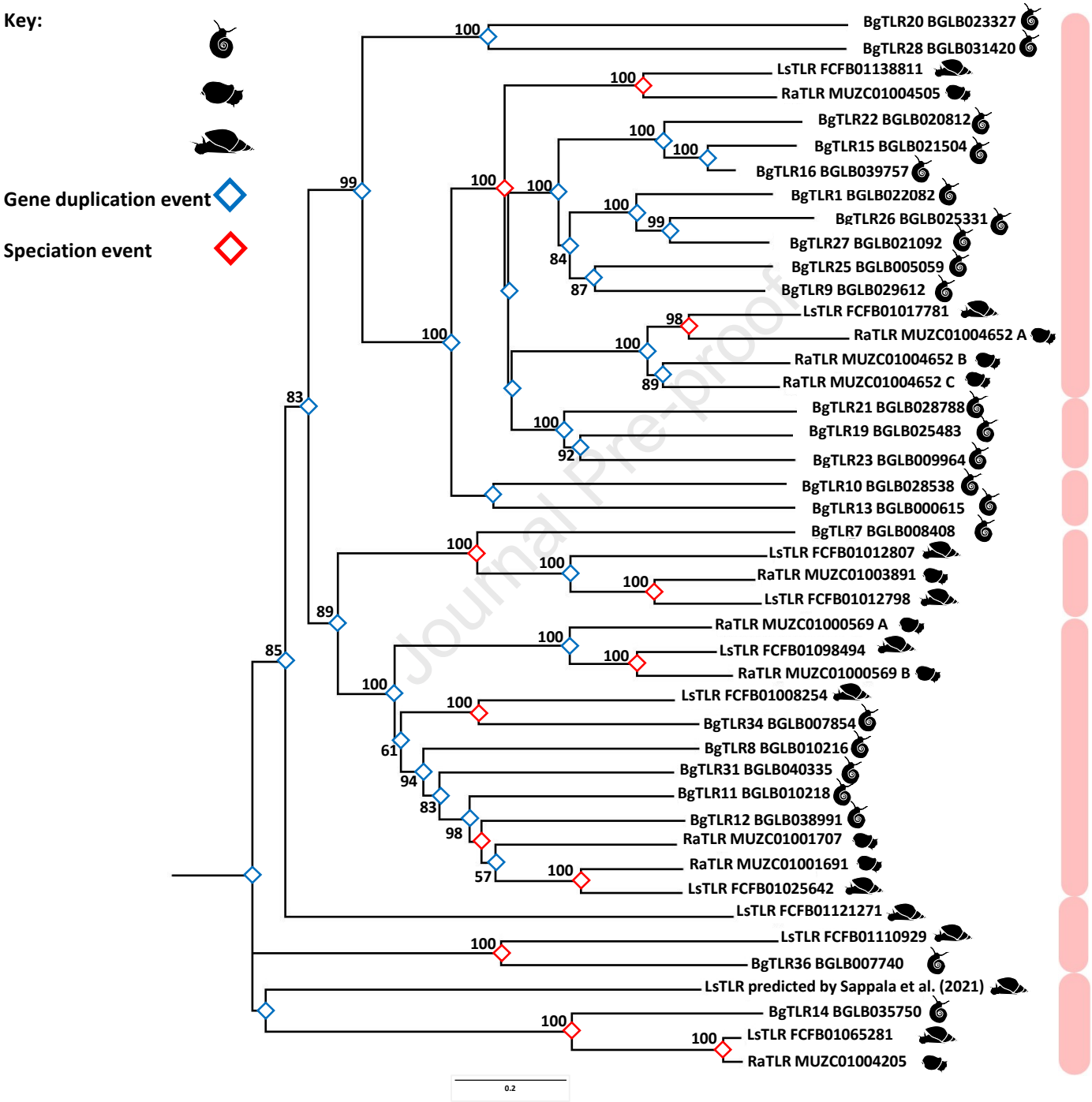
B) *Lymnaea stagnalis*

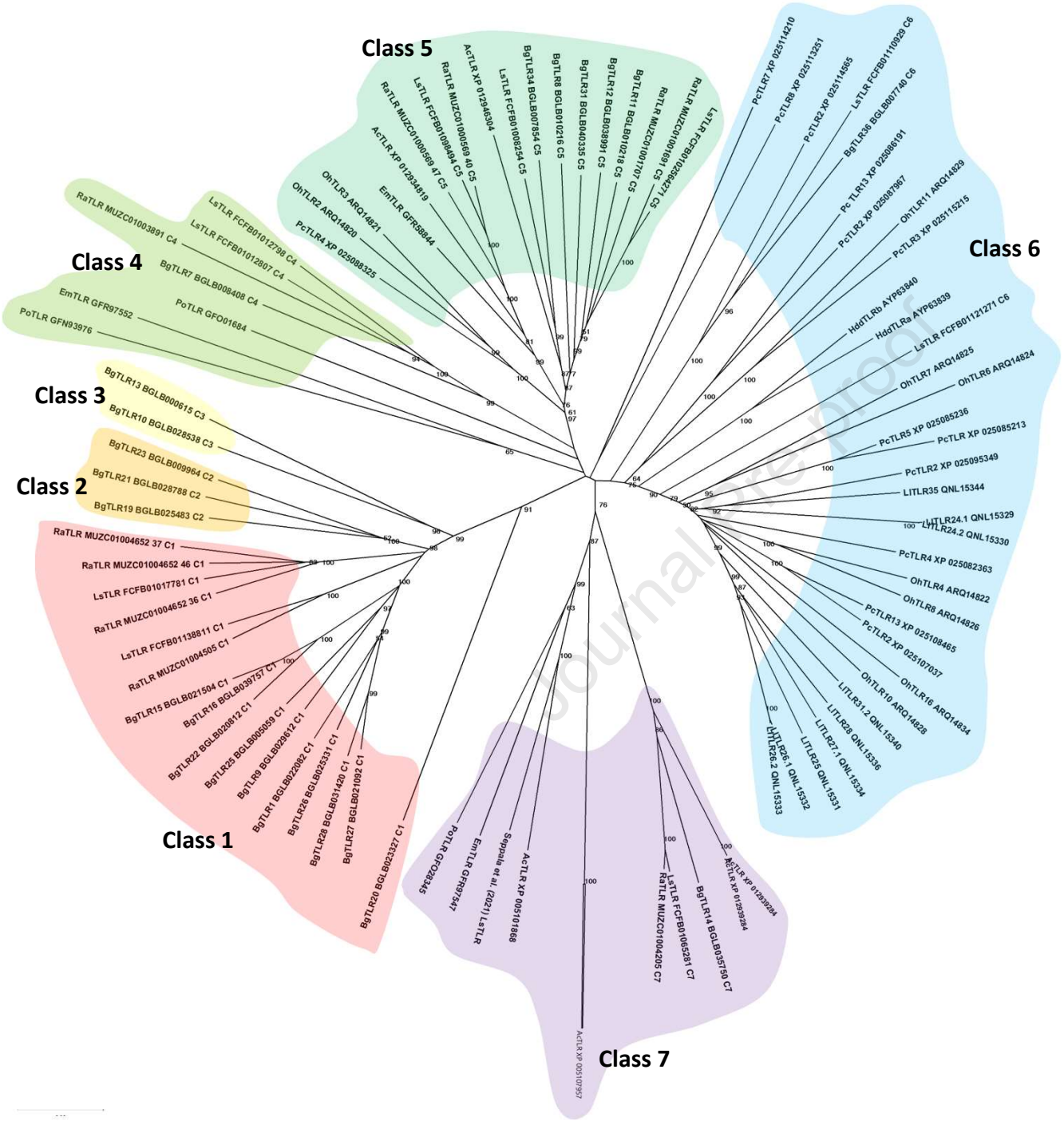
i)



ii)







Key: Gastrpod toll li e receptor (TLR) no enclature

BgTLR: ***Biomphalaria glabrata***

Ra TLR: ***Radix auricularia***

LsTLR: ***Lymnaea stagnalis***

TLR: ***Oncomelania hupensis***

cTLR: ***Pomacea canaliculate***

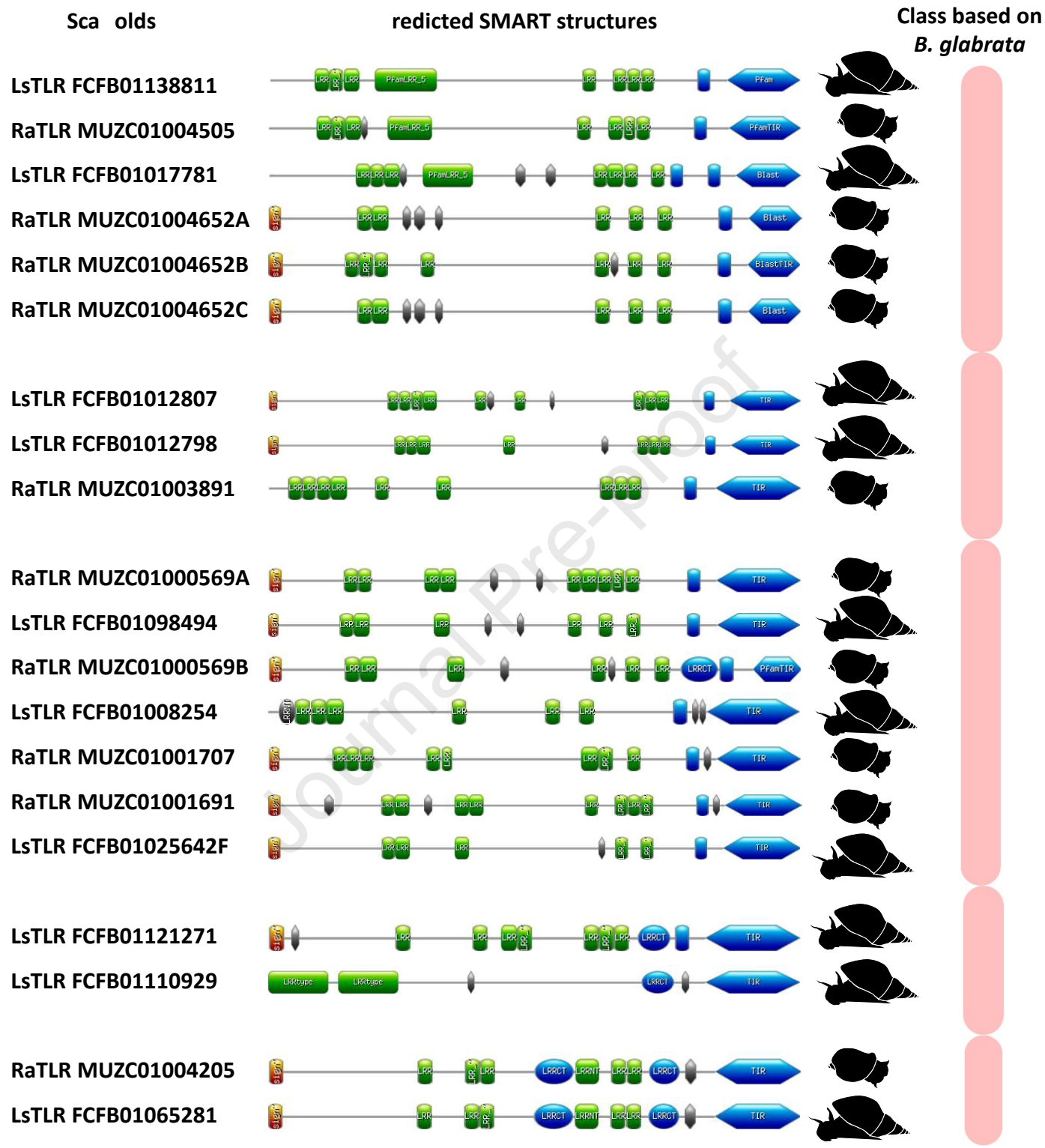
LITLR: ***Littorina littorea***

ddTLR: ***Haliotis discus discus***

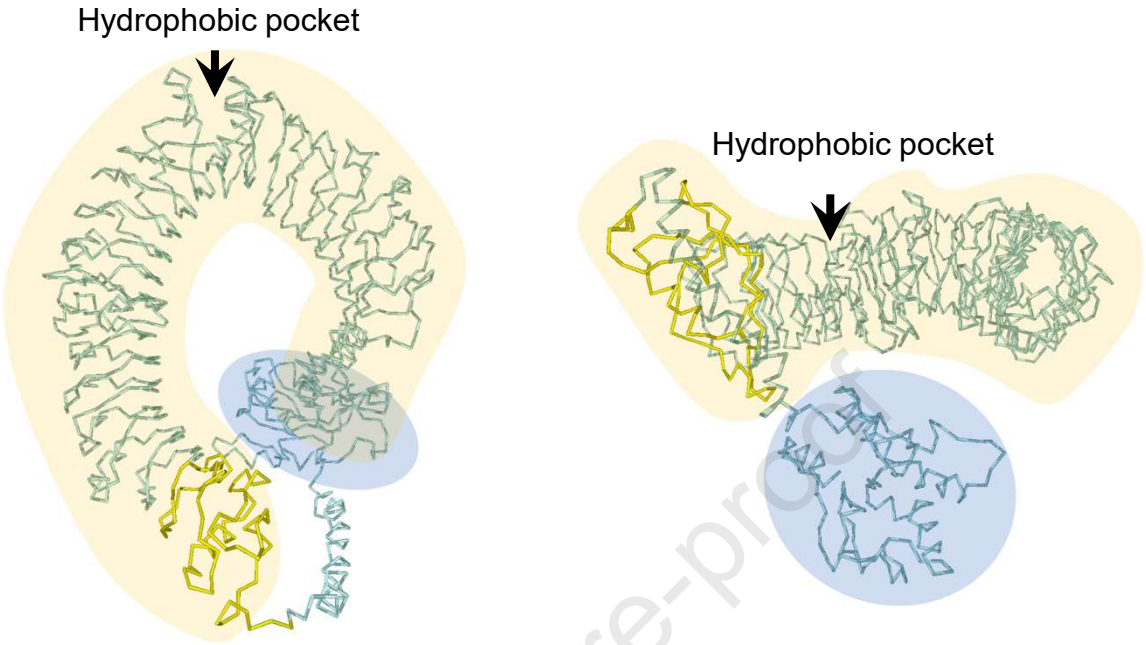
TLR: ***Elysia marginate***

AcTLR: ***Aplysia californica***

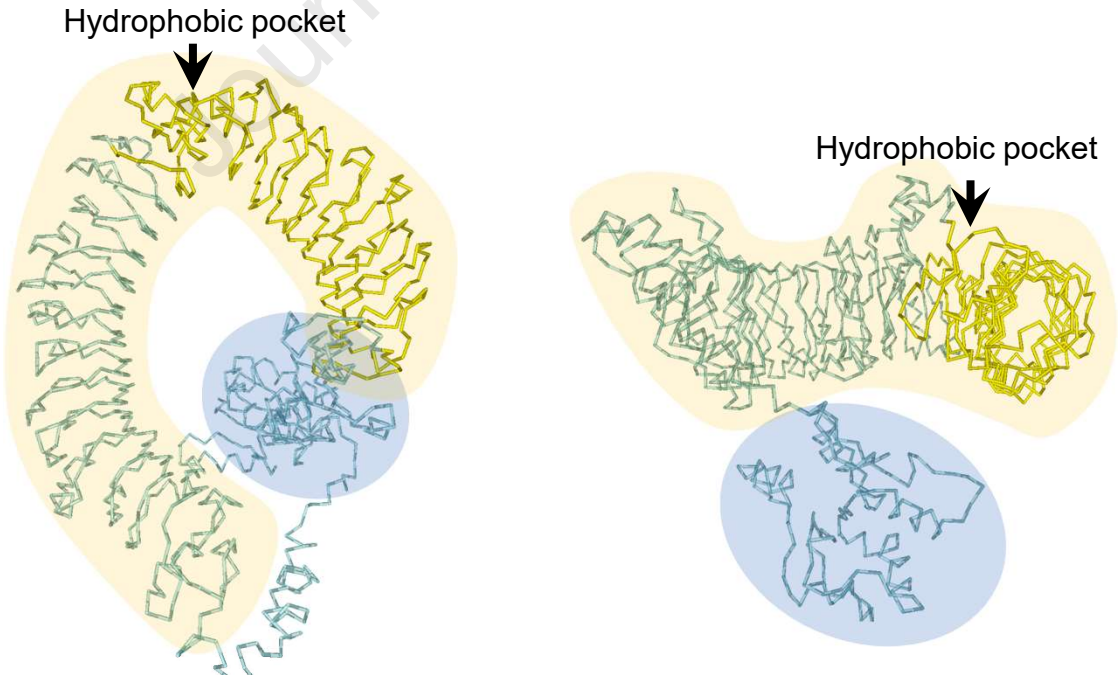
oTLR: ***Plakobranchus ocellatus***



A) **Class 1 LsTLR FCFB01138811**



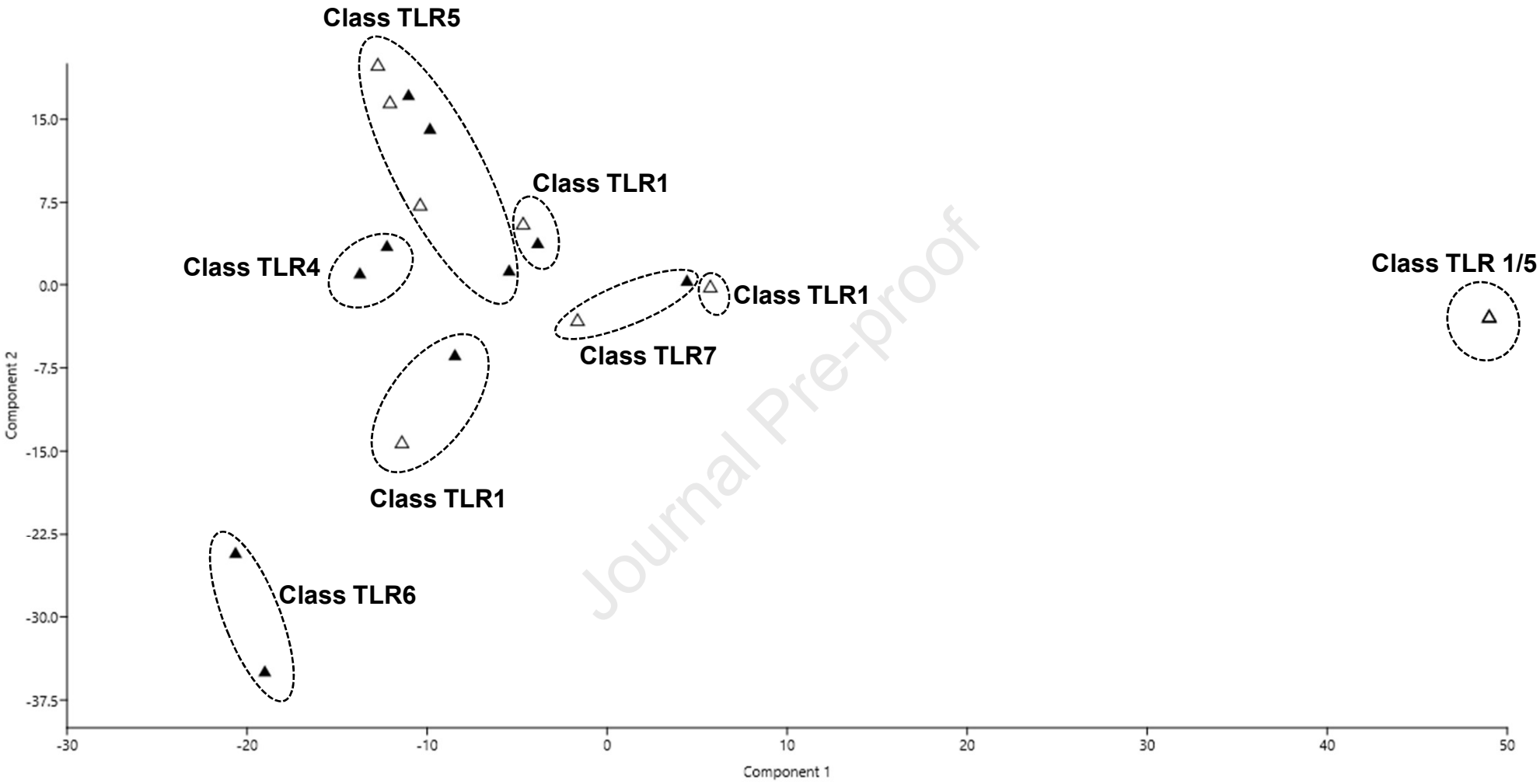
B) **Class 1 RaTLR MUZC01004652 C**



Key:

LRR Region

TIR Domain



Key:

Radix auricularia △

Lymnaea stagnalis ▲

Highlights:

- First attempt to predict and analyse putative TLR proteins in lymnaeid species
- Indentification of the occurance of molluscan class 1 TLRs in more species not only *Biomphalaria glabrata*
- Identification of both single cysteine clusters and multiple cysteine clusters TLRs in lymnaeids
- Gene duplication and structural radiation events of TLRs in lymnaeids could be indicative of adapation to a multiple pathogen environment