THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

An expressed sequence tag survey of gene expression in the pond snail Lymnaea stagnalis, an intermediate vector of Fasciola hepatica

Citation for published version:
Davison, A & Blaxter, ML 2005, 'An expressed sequence tag survey of gene expression in the pond snail Lymnaea stagnalis, an intermediate vector of Fasciola hepatica', *Parasitology*, vol. 130, no. 5, pp. 539-552. https://doi.org/10.1017/S0031182004006791

Digital Object Identifier (DOI):
10.1017/S0031182004006791

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Parasitology

Publisher Rights Statement:
RoMEO green

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 03. May. 2024

An expressed sequence tag survey of gene expression in the pond snail *Lymnaea stagnalis*, an intermediate vector of *Fasciola hepatica*

A. DAVISON[1,2]* *and* M. L. BLAXTER[1]

[1] *School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, UK*
[2] *Institute of Genetics, School of Biology, University of Nottingham, Nottingham NG7 2UH, UK*

(*Received 2 August 2004; revised 23 September 2004; accepted 28 September 2004*)

SUMMARY

The pond snail *Lymnaea stagnalis* is an intermediate vector for the liver fluke *Fasciola hepatica*, a common parasite of ruminants and humans. Yet, despite being a disease of medical and economic importance, as well as a potentially useful comparative tool, the genetics of the relationship between *Lymnaea* and *Fasciola* has barely been investigated. As a complement to forthcoming *F. hepatica* expressed sequence tags (ESTs), we generated 1320 ESTs from *L. stagnalis* central nervous system (CNS) libraries. We estimate that these sequences derive from 771 different genes, of which 374 showed significant similarity to proteins in public databases, and 169 were similar to ESTs from the snail vector *Biomphalaria glabrata*. These *L. stagnalis* ESTs will provide insight into the function of the snail CNS, as well as the molecular components of behaviour and response to parasitism. In the future, the comparative analysis of *Lymnaea*/*Fasciola* with *Biomphalaria*/*Schistosoma* will help to understand both conserved and divergent aspects of the host-parasite relationship. The *L. stagnalis* ESTs will also assist gene prediction in the forthcoming *B. glabrata* genome sequence. The dataset is available for searching on the world-wide web at http://zeldia.cap.ed.ac.uk/mollusca.html.

Key words: *Biomphalaria*, expressed sequence tag, *Fasciola*, *Lymnaea*, neuropeptide, snail.

INTRODUCTION

Although fascioliasis is often only considered to be a disease of economic importance, up to 17 million people are infected world-wide (Esteban, Bargues & Mas-Coma, 1998; Hurtrez-Bousses *et al*. 2001). The Lymnaeid snails are the principal intermediate vectors for the disease, caused by the liver fluke *Fasciola hepatica* (Hurtrez-Bousses *et al*. 2001). Snails become infected with *F. hepatica* when a motile miracidium hatches from parasite eggs in faeces-contaminated water and penetrates their body. The cycle is completed when cercaria emerge from the snail, attach to submerged vegetation and are eaten by the definitive host, most frequently a ruminant. The life-cycle of *F. hepatica* in *Lymnaea stagnalis* is typical of the dignean platyhelminthes and similar to that of schistosomes in *Biomphalaria glabrata*, suggesting that a comparison between different systems may help understand commonalities in the host-parasite relationship (Bayne, Hahn & Bender, 2001; Sorensen & Minchella, 2001).

The *Lymnaea* – *Fasciola* pairing has been investigated with respect to life-history, prevalence, and population dynamics of the host-parasite interaction, but to a much lesser extent than has *Biomphalaria* and its *Schistosoma* parasites (Hurtrez-Bousses *et al*. 2001; Mas-Coma, Funatsu & Bargues, 2001; Sorensen & Minchella, 2001). Fasciolid genetics has rarely been investigated (Vignoles, Dreyfuss & Rondelaud, 2002; Meunier *et al*. 2004). As of April 2004, only 140 *F. hepatica* and 165 *L. stagnalis* DNA sequences had been submitted to GenBank. In comparison, the *S. mansoni* genome project is approaching completion (http://www.sanger.ac.uk/Projects/S_mansoni/), and *F. hepatica* ESTs will be produced shortly, following an initiative at the Sanger Centre (http://www.sanger.ac.uk/Projects/S_mansoni/). Large-scale sequencing of *B. glabrata* expressed sequence tags (ESTs) is under way, a BAC library is available, and the genome will shortly be sequenced to 4 to 6-fold coverage by the National Human Genome Research Institute (http://biology.unm.edu/biomphalaria-genome/; http://www.genome.gov/11007951).

While the outcome of an attack by a schistosome parasite is genetically determined (Bayne *et al*. 2001; Jones *et al*. 2001), genetic factors that control the resistance to trematode infection are relatively poorly understood in both *L. stagnalis* (Hoek *et al*. 1997) and the definitive hosts (Piedrafita *et al*. 2004). In contrast, an understanding of the genetic basis of the host-parasite relationship is more advanced in *B. glabrata*, where candidate genes involved in parasite clearance (Knight *et al*. 1998, 1999; Jones

* Corresponding author: Institute of Genetics, University of Nottingham, Queens Medical Centre, Nottingham NG7 2UH, UK. Tel: +44 (0)115 8493239. Fax: +44 (0)115 9709906. E-mail: angus.davison@nott.ac.uk

 © 2004 Cambridge University Press
doi:10.1017/S0031182004006791 Printed in the United Kingdom

et al. 2001), and genes that are upregulated during parasitosis (Léonard *et al*. 2001; Miller *et al*. 2001; Raghavan *et al*. 2003; Nowak *et al*. 2004) have been isolated (reviewed most recently by Knight, Ongele & Lewis, 2000; Lockyer *et al*. 2004). Ultimately, only detailed studies of snail genes that are expressed upon exposure to a parasite will improve our understanding of snail resistance to infection.

L. stagnalis has been used as a model in studies of neuronal function (Smit *et al*. 1992, 2001; Smit, Hoek & Geraerts, 1993), toxicology (Frantsevich *et al*. 1996; Salanki, 2000), the molecular basis of behaviour and learning (Chase, 2002; Lukowiak *et al*. 2003), and the evolution of body asymmetry (Hosoiri, Harada & Kuroda, 2003). Snails in general are useful for understanding how natural selection operates (Davison, 2002). Thus, the present study was initiated with the aim to rapidly identify genes in *L. stagnalis*, and compare them with genes in *B. glabrata* and other organisms, through EST analysis. In this approach, DNA sequences derived from randomly selected cDNAs can be used to define the genes expressed by an organism. It is a useful technique where no prior sequence information is available (Daub *et al*. 2000; Kenyon *et al*. 2003). As a first step, two *L. stagnalis* central nervous system (CNS) cDNA libraries were used, because it has been reported previously that neuropeptide genes moderate behaviour during parasitosis (Hoek *et al*. 1997), but also because *L. stagnalis* is a model for neuronal function.

MATERIALS AND METHODS

cDNA library preparation and expressed sequence tag generation

Two lambda cDNA libraries were prepared previously using similar methods, both described in detail by Hoek *et al*. (1997) and Sadamoto *et al*. (2004). Both lab stocks derive from the same wild collection (circa 1960), from Eempolder, The Netherlands. The libraries were plated on *E. coli XL-1*-Blue cells and recombinant clones picked at random. The cDNA inserts of clones were amplified by PCR using the universal T3 and T7PL primers and sized on a 1·2% agarose gel. PCR products were cleaned and sequenced using the SAC primer (GGGAACAAAAGCTGGAG) with Big Dye v3.0 and an ABI 3730 automated sequencer. The phage stocks were archived at −70 °C, and are available for research purposes on request. Several clones were also completely sequenced, using an internal, clone-specific sequencing primer.

Bioinformatics

Vector and poor quality sequences were removed using an automated method, implemented in TRACE2DBEST, then clustered on the basis of sequence identity using CLOBB, both programs within PARTIGENE (Parkinson *et al*. 2004; available from http://www.nematodes.org/PartiGene/). Briefly, CLOBB is an iterative clustering method where sequences are grouped together on the basis of BLAST similarity. The program identifies 'superclusters' of related clusters and attempts to avoid expansion of chimeric clusters. A consensus sequence was derived for each cluster. Each consensus sequence was compared to the public databases (GenBank non-redundant nucleotide and protein databases and *B. glabrata* ESTs) using the BLAST algorithms (Altschul *et al*. 1990, 1997) within PARTIGENE. Genes were assigned to gene classes, first by translating them using prot4EST v2.0 (James Wasmuth, unpublished; downloaded from http://www.nematodes.org/PartiGene/), then gene ontology annotating them using GOblet (Hennig, Groth & Lehrach, 2003). The output clusters (e.g. LSC00010) and analyses are available at http://zeldia.cap.ed.ac.uk/mollusca.html. The website also has an equivalent cluster analysis for *B. glabrata*, and each cluster is fully searchable, by keyword and BLAST.

Prediction of signal peptides and cleavage sites was carried out using SignalP 3.0 (Bendtsen *et al*. 2004) and ProP 1.0 (Duckert, Brunak & Blom, 2004). Sequence conservation plots were produced using the Sequence Manipulation Suite 2 (http://bioinformatics.org/sms2/), using pre-defined groups of similar amino acids (GAVLI, FYW, CM, ST, KRH, DENQ, P). In the plots, positions with identical amino acids (above a threshold proportion) are shaded black, whereas positions with different, but same-group amino acids are shaded grey.

Phylogenetics

Phylogenetic analyses were performed using PAUP* (version 4.0b10). Evolutionary trees were constructed using the neighbour-joining method (NJ) with distances corrected for multiple hits by using the general time-reversible (GTR) model with between-site rate heterogeneity accounted for by incorporating a proportion of invariant sites and gamma-distributed rates into the model. The rate matrix, base frequencies, proportion of invariant sites and shape parameter (alpha) of the gamma distribution (based on 16 rate categories) were estimated using likelihood by iteration from an initial neighbour-joining tree. The parameters estimated from the initial tree were then used to build a new neighbour-joining tree and the parameters re-estimated. This process was repeated until there was no further improvement in likelihood.

RESULTS

The lambda inserts of the cDNA library developed by Hoek *et al*. (1997) were randomized with respect

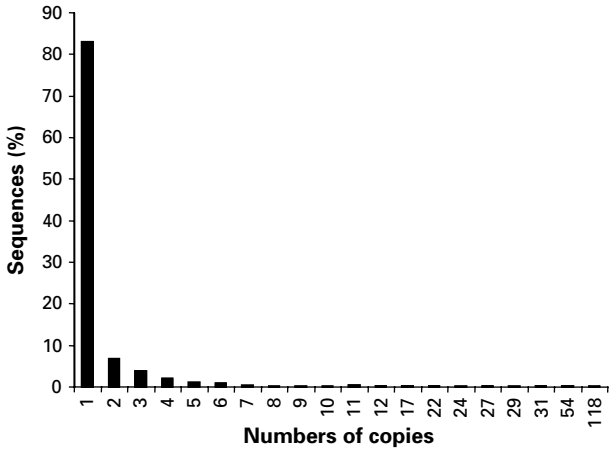

Fig. 1. Assessment of redundancy of the *Lymnaea stagnalis* EST dataset. The ESTs were clustered into putative genes. The graph shows the relative abundance of each EST-size-class of cluster. Most genes were represented by a single EST.

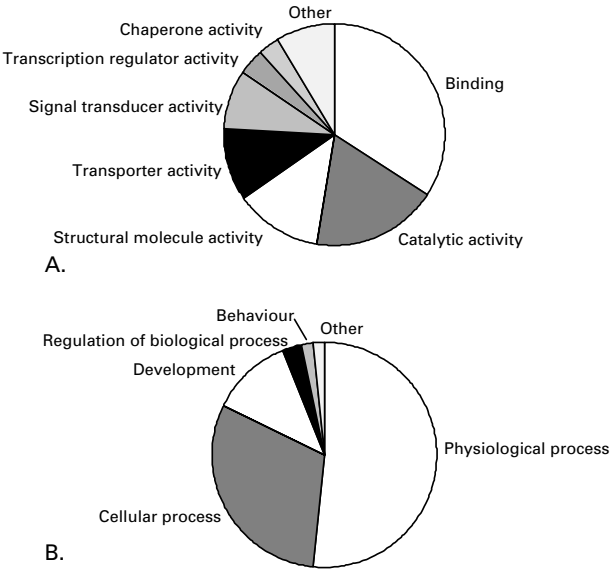

Fig. 2. Annotated GO terms and protein domains in the *Lymnaea stagnalis* EST dataset. (A) GO Molecular Function (GO:0003674; 261 matches). (B) GO Biological Process (GO:0008150; 252 matches). In the Cellular Process subcategory (GO:0009987), the largest groups were associated with cell communication, cell growth and/or maintenance and cell death. In the Physiological Process subcategory (GO:0007582), the largest groups were associated with metabolism, stimulus response, and neurophysiological processes.

to orientation. This made single pass sequencing less efficient (only half of all sequencing reactions work due to problems sequencing through the poly-A tail), so only 33 sequences from this library were used. The remainder of the sequences (1287) were from the library developed by Sadamoto *et al*. (2004). From a total of 1320 sequences longer than 100 bp (GenBank Accession numbers CN809706-CN811025), the analysis yielded 771 clusters (see http://zeldia.cap.ed.ac.uk/mollusca.html), with 374 (48%) showing significant BLASTx similarity to another sequence in the nonredundant databases. 650 EST clusters (49%) had one sequence only, and 253 (39%) of these singletons had significant homology to another sequence in the databases (Fig. 1).

In total, 276 *L. stagnalis* clusters were assigned Gene Ontology (GO) terms using GOblet, and higher-level GO terms were extracted for the 3 GO domains (Fig. 2). Of these, 261 had terms for molecular function assigned, and within these 'binding' and 'catalytic activity' were the most common GO assignments (Fig. 2A). Of the clusters with GO molecular function terms 12% were assigned to 'transporter activity' or 'signal transduction', including ion transporters, protein transporters and transmembrane neuroreceptors. In the GO biological process domain (Fig. 2B), just over half of the assignments were to 'physiological processes' such as metabolism (including neurotransmitter and amine metabolism), stimulus response, and secretion of neurotransmitters.

Previous to this work, genes reported previously in *L. stagnalis* (165 DNA sequences, 304 proteins; GenBank May 2004) include high copy-number genes isolated for phylogenetic analyses (Remigio & Blair, 1997; Remigio & Hebert, 2003), and high abundance peptides expressed in the central nervous system (e.g. Smit *et al*. 1992, 1993, 2001; Kellett *et al*. 1994; Table 1). Only a small proportion of the ESTs (~9%) were derived from previously isolated *L. stagnalis* sequences, but 12 of the most abundant ESTs that we isolated have been isolated previously as cDNAs (Table 1). We also identified neuropeptides novel to *L. stagnalis* (Fig. 2), an immune defence gene (Fig. 3), and an abundant gene family with features suggestive of secretion and processing to form smaller peptide signal molecules (Fig. 4; 8 peptides were derived from 10 different DNA sequences).

As the public sequence databases are growing exponentially, BLAST-based analysis of similarities can only represent a snapshot in time of the relationships and putative functions of genes identified through EST sequencing. The *B. glabrata* EST set is also growing rapidly, but in May 2004, 169 or 22% of the *L. stagnalis* cluster consensus sequences had a significant match to a *B. glabrata* sequence. Abundant sequences in *L. stagnalis* are also more likely to have been isolated in *B. glabrata*, because clusters of two or more sequences more often have a *B. glabrata* sequence match (37%) compared with singleton clusters (16%). Some of these similarities to *B. glabrata* are shown in Table 2.

ESTs are an effective way of sampling the expressed genome of an organism, and are a route to rapid identification of conserved genes in otherwise neglected taxa. Thus, many of the *L. stagnalis* ESTs are derived from genes that are probably universally

Table 1. Genes expressed abundantly in the *Lymnaea stagnalis* central nervous system

(Many of the abundant clusters have been identified previously in *L. stagnalis* or a related mollusc. In contrast, ~52% of the other sequences showed no significant similarity to other proteins.)

No. of ESTS	Blast e-value[1]	Putative gene identity	Putative function	Orthologue in	Notes	Cluster No.[3]	GenBank[4]	Reference
118		16S ribosomal RNA	mitochondrial gene	*L. stagnalis*		LSC00004	CN809741	Remigio & Blair (1997)
54		ovulation prohormone precursor	ovulation hormone	*L. stagnalis*	5 contigs	LSC00001	CN809749	Ebberink *et al.* (1985)
31	1E-63	APGWamide	neuropeptide	*L. stagnalis*		LSC00018	CN809782	Smit *et al.* (1992)
31		FMRFamide-related neuropeptide precursor	neuropeptide	*L. stagnalis*		LSC00190	CN809795	Kellett *et al.* (1994)
27	no significant hits[2]	not known			8 contigs	LSC00059 LSC00481	CN809786	
25		18S ribosomal RNA	nuclear ribosomal gene	*L. stagnalis*		LSC00046	CN809768	Bargues *et al.* (2003)
24		acetylcholine-binding protein precursor	modulator of synaptic transmission	*L. stagnalis*		LSC00044	CN809766	Smit *et al.* (2001)
12		soma ferritin	iron storage	*L. stagnalis*		LSC00113	CN809874	von Darl, Harrison & Bottke (1994)
12	3E-174	beta tubulin 1 and 2	structural	*Octopus vulgaris*	4 contigs	LSC00313	CN809824	Noto & Endoh (2004)
11		cytochrome c oxidase subunit I	mitochondrial gene	*L. stagnalis*		LSC00003	CN809745	Remigio & Hebert (2003)
11		preproLYCP	neuropeptide	*L. stagnalis*		LSC00041	CN809763	Smit, Hoek & Geraerts (1993)
11	3E-18	glial-specific Ag protein precursor	neuropeptide	*A. californica*		LSC00043	CN809765	Lockhart, Levitan & Pikielny (1996)
21		pedal peptide preprohormone	neuropeptide	*L. stagnalis*		LSC00129	CN809872	Hoek *et al.* (1997)
19		cytochrome P450	mono-oxygenase	*L. stagnalis*		LSC00009	CN809785	Teunissen *et al.* (1992)
14	1E-100	insulin related peptide	possible neurotransmitter	*L. stagnalis*		LSC00267	CN810093	Smit *et al.* (1996)
10		alpha tubulin 1 and 2	structural	*A. californica*	2 contigs	LSC00084	CN809824	Moccia *et al.* (2003)
8	1E-144	elongation factor 1-alpha	protein biosynthesis	*B. glabrata*		LSC00016	CN809860	Lockyer (unpublished)
8	3E-79	calmodulin	signalling	*Patinopecten sp.*		LSC00053	CN809778	Toda *et al.* (1981)
7	no significant hits	not known				LSC00091	CN809834	
6	1E-13	sensorin A precursor	neuropeptide	*A. californica*		LSC00533	CN810462	Brunet *et al.* (1991)
6	2E-123	actin	structural protein	*B. glabrata*	2 contigs	LSC00025	CN809740	Adema (2002)
5	no significant hits	not known				LSC00128	CN809896	
5	no significant hits	not known				LSC00130	CN809898	
6	6E-54	myoglobin	oxygen transport	*B. glabrata*		LSC00360	CN810207	Dewilde *et al.* (1998)
5		28s ribosomal RNA	nuclear ribosomal gene	*Arion silvaticus*		LSC00088	CN809829	Passamaneck *et al.* (unpublished)
5	no significant hits	not known				LSC00076	CN809810	
5	no significant hits	not known				LSC00079	CN809814	
5	no significant hits	not known			2 contigs	LSC00185	CN809978	
5	no significant hits	not known				LSC00239	CN810055	
5	1E-56	cytochrome c oxidase subunit II	mitochondrial gene	*B. glabrata*		LSC00348	CN810811	Dejong, Emery & Adema (2004)
5		poly ubiquitin C	proteolytic targeting	*B. glabrata*		LSC00159	CN809938	Wu & Yoshino (unpublished)

[1] e=0 unless stated.
[2] see Figure 4.
[3] See zeldia.cap.ed.ac.uk/mollusca.html.
[4] dbEST reference for first occurence of sequence.

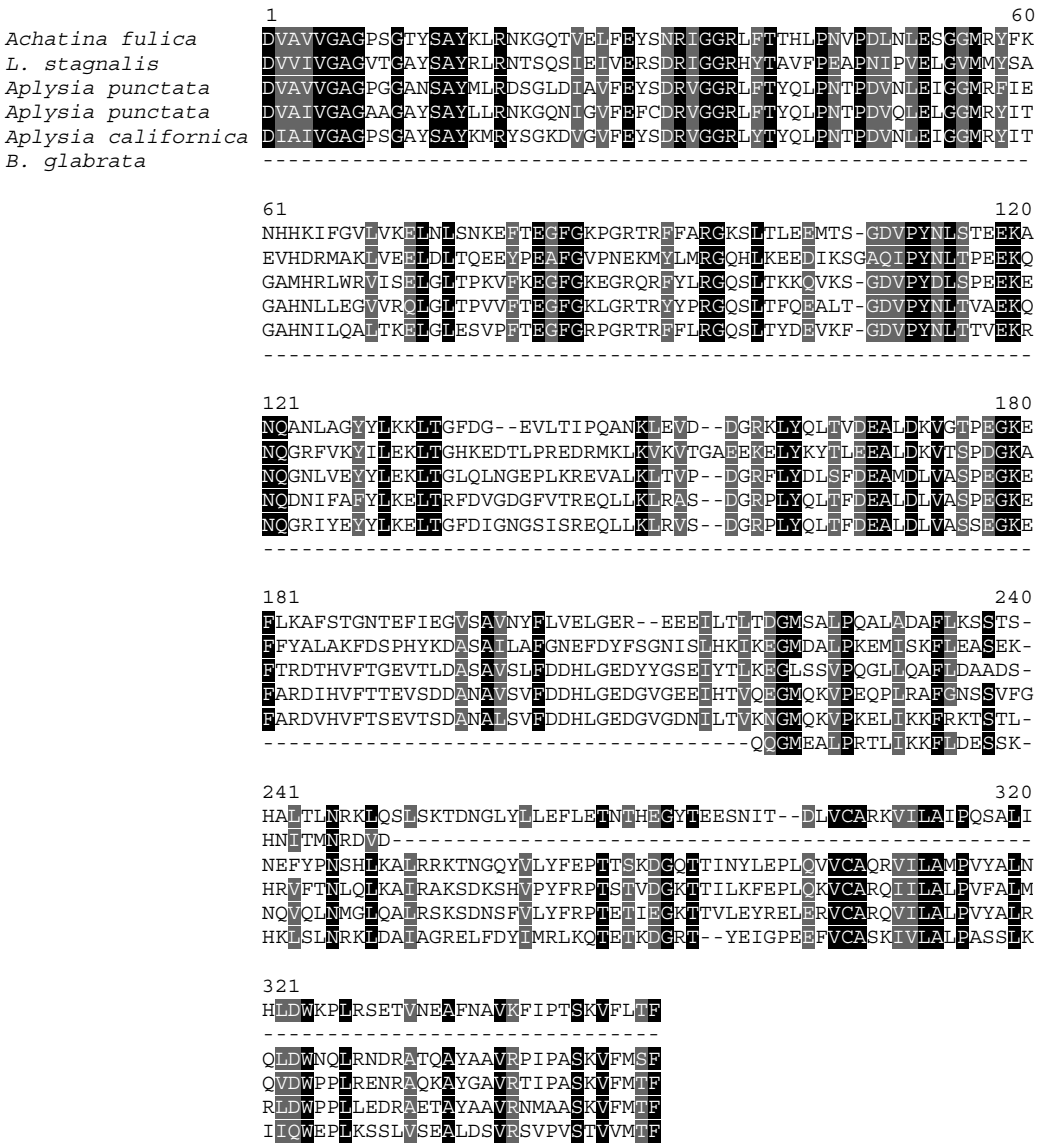

Fig. 3. Conservation plot of *achacin*-like partial protein sequences from *Aplysia punctata* (AY442281, AAR14186, AAR14187, CAC19361, CAC19362), *Aplysia californica* (AAN78211), *Achatina fulica* (P35903), *Biomphalaria glabrata* (CN476061, CN810215, CN810515, CN810525, CN810863) and *Lymnaea stagnalis* (CN476059, CN445880, CN445878, CN476093; cluster LSC00366). Both the *L. stagnalis* and *B. glabrata* sequences are partial.

present in animals, but had not previously been reported from the superphylum Lophotrochozoa, of which Mollusca is a member (Table 3). These genes may be useful for deeper phylogenetic analyses (e.g. Elongation factor 1 alpha, clusters LSC00016, LSC00124), as will the discovery of 7 of the expected 13 mitochondrial genome-encoded genes (Table 4).

DISCUSSION

The announcement of the *B. glabrata* genome project will provide an invaluable resource to facilitate gene discovery and genetic mapping in *Biomphalaria* and other snails. Complementary to that, the *L. stagnalis* ESTs that we have isolated will facilitate gene prediction in the complete *Biomphalaria* genome, because although many (~78%) do not have current BLAST hits to a *Biomphalaria* EST, we expect that most of these genes are conserved between these two relatively closely related species. However, out of necessity, we here concentrate our analysis on genes that have been isolated in both species or other species (~48% of the total), especially those which are interesting from a parasitological perspective.

Lymnaea *orthologues of* Biomphalaria *resistance genes*

There has been an extensive and rapidly increasing effort to elucidate the genes involved in conferring resistance of *Biomphalaria* to infection by schistosomes and other parasites, work which was reviewed most recently by Knight *et al*. (2000) and Lockyer

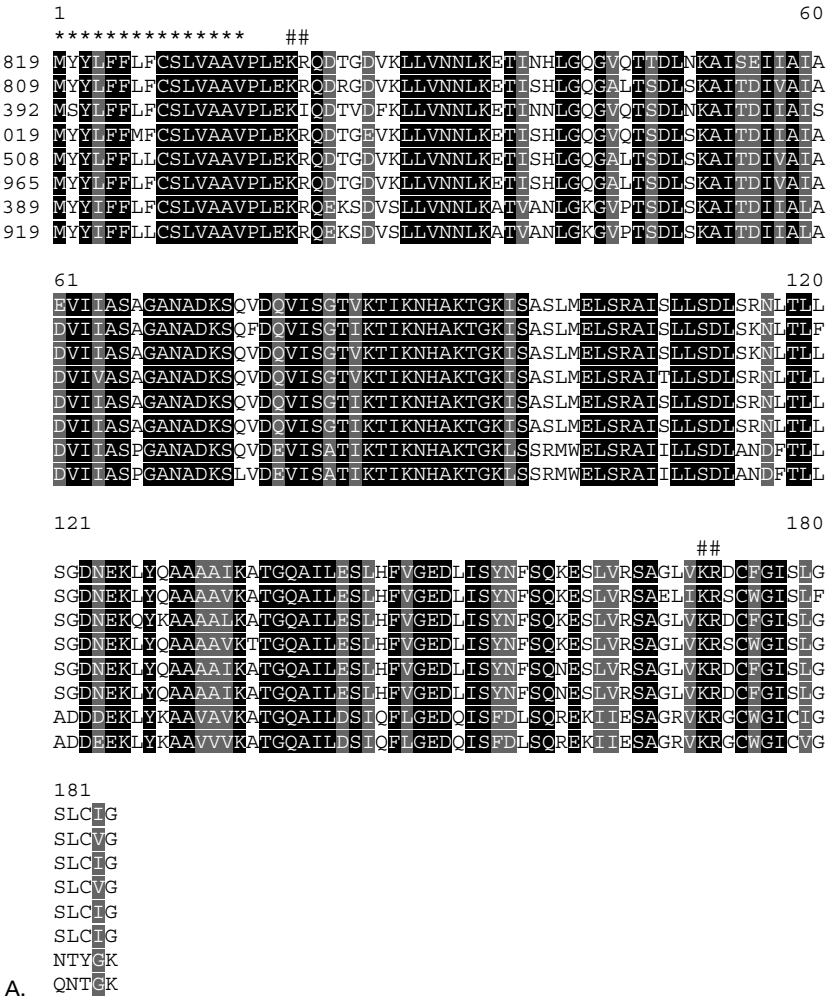

Fig. 4. A putative novel gene family in *Lymnaea stagnalis* identified by EST analysis (clusters LSC00481 and LSC00059). (A) Sequence conservation plot of the predicted near full length translations of 8 different peptides, including the signal region (*). Two putative lysine (K)/arginine (R) propeptide cleavage sites are present (#). (B) Neighbour-joining DNA phylogeny of the gene family. Two divergent groups were recovered, with one containing 4 separate groups. Bootstrap support $>95\%$ is shown by an asterisk.

Table 2. Selected genes isolated from both *Lymnaea stagnalis* and *Biomphalaria glabrata*

Cluster No.[1]	*Lymnaea* GenBank[2]	Putative gene identity	Blast e-value (*L.s*:GenBank)	*Biomphalaria* GenBank	Blast e-value (*L.s.*:*B.g.*)
LSC00159	CN809938	Polyubiquitin c	1E-159	AF486292	0
LSC00023	CN810038	Eukaryotic translation elongation factor 2	1E-134	CK149257	1E-146
LSC00290	CN810118	Thioredoxin peroxidase	1E-119	AY026258	1E-145
LSC00296	CN810619	Ribosomal protein S2	1E-116	CK988682	1E-100
LSC00106	CN809859	Ribosomal protein S4	1E-114	AW739514	5E-93
LSC00653	CN810636	Robosomal protein S6	1E-101	CK989088	1E-124
LSC00496	CN810411	Phosphoglycerate kinase	5E-86	CK989213	3E-77
LSC00220	CN810025	Ras-related nuclear protein	2E-83	CK149192	1E-109
LSC00016	CN810781	Elongation factor 1-alpha	1E-82	CK989121	1E-106
LSC00740	CN810792	Enolase	5E-82	CK990028	1E-61
LSC00237	CN810161	Ribosomal protein S3a	3E-80	CK989183	1E-101
LSC00025	CN810640	Actin	1E-79	CK989669	1E-99
LSC00053	CN810943	Calmodulin	2E-79	CK988750	7E-78
LSC00101	CN809852	Ribosomal protein L18	2E-79	CK988753	5E-88
LSC00269	CN810095	High mobility group protein	2E-77	AF484963	2E-95
LSC00773	CN810844	Paramyosin	4E-77	CK656810	1E-112
LSC00252	CN810171	Cytoplasmic intermediate filament protein A	2E-70	CK149507	2E-79
LSC00561	CN810500	H3 histone	6E-70	CK989054	3E-76
LSC00099	CN810024	Heat shock protein 70	4E-69	AFO25477	4E-71
LSC00811	CN810898	Ribosomal protein L9	1E-63	CK989044	1E-104
LSC00113	CN809874	Ferritin heavy chain	6E-63	CK988653	9E-74
LSC00563	CN810503	Ribosomal protein L23	2E-62	CK990055	4E-82
LSC00299	CN810129	Ribosomal protein L13	1E-61	CK989942	2E-46
LSC00341	CN810184	Ribosomal protein L32	2E-58	CK988843	1E-63
LSC00029	CN810373	Peptidyl-prolyl cis-trans isomerase	2E-54	CK989543	4E-61
LSC00350	CN810391	Ribosomal protein L10a	4E-54	CK990128	5E-69
LSC00233	CN810666	Y-box factor	6E-52	CK988763	1E-37
LSC00871	CN810987	Glycosyl hydrolase family 9 endoglucanase	2E-51	CK149333	2E-47
LSC00794	CN810875	Moesin/ezrin/radixin	2E-47	AF379610	6E-69
LSC00187	CN809980	Ribosomal protein L14	1E-42	CK989458	6E-74
LSC00398	CN810262	Myosin II	5E-41	CK989977	1E-72
LSC00349	CN810193	Ribosomal protein S15	1E-40	AW739697	2E-68
LSC00538	CN810469	Tropomyosin	5E-40	M97554	1E-106
LSC00366	CN810215	Achacin	1E-39	CN476061	1E-05
LSC00631	CN810603	Ribosomal protein L21	3E-37	CK989043	9E-46
LSC00288	CN810582	Cartilage matrix protein precursor (matrilin-1)	6E-37	AF486289	1E-43
LSC00243	CN810187	Nucleolar protein family A member 2	8E-34	CD760658	1E-46
LSC00336	CN810179	Bactericidal permeability increasing protein	2E-31	CK988821	4E-40
LSC00209	CN810641	Translationally-controlled tumor protein	6E-31	CK988694	2E-97
LSC00502	CN810418	Ribosomal protein S13	7E-30	CK989888	3E-42
LSC00686	CN810706	Ribosomal protein L7	3E-26	CN445825	1E-43
LSC00048	CN810456	ATP synthase c-subunit	7E-19	CK988966	9E-57

[1] zeldia.cap.ed.ac.uk/mollusca.html.
[2] dbEST reference for first occurence of sequence.

et al. (2004), and described extensively in a special issue of *Parasitology* (vol. 123, issue 7). While an in depth review of the molecular genetics of *B. glabrata*/*S. mansoni* interaction is beyond the scope of this discussion, we were interested to know if there were any orthologues of the putative resistance genes (or genes linked to resistance genes) in our dataset, even though the snails had not been exposed to parasites.

The response to parasitism is poorly understood in *L. stagnalis*, and has been studied in greater depth at the molecular genetic level in *B. glabrata* (Knight *et al*. 1998, 1999; Jones *et al*. 2001; Léonard *et al*. 2001; Miller *et al*. 2001; Raghavan *et al*. 2003; Nowak *et al*. 2004). In both species, the main line of defence is the circulating haemocyte cell type but, at least in *B. glabrata*, a range of soluble factors such as

Table 3. Genes isolated from *Lymnaea stagnalis* that have not been previously described in the Lophotrochozoa

Blast e-value	Putative gene identity	Cluster No.[1]	GenBank[2]
1E-118	pre-mRNA splicing factor RNA helicase	LSC00595	CN810558
1E-114	RuvB DNA helicase	LSC00575	CN810526
1E-111	DEAD box protein	LSC00711	CN810746
1E-103	ADP, ATP carrier protein	LSC00599	CN810562
1E-102	60S ribosomal protein L10	LSC00640	CN810617
1E-93	Rab GDP dissociation inhibitor	LSC00352	CN810197
2E-91	Alpha coat protein	LSC00217	CN810019
1E-88	40S ribosomal protein S9	LSC00144	CN809914
5E-85	Arginosuccinate synthetase	LSC00700	CN810732
4E-84	60S acidic ribosomal protein	LSC00860	CN810971
1E-81	Dynein	LSC00449	CN810343
8E-78	26S proteasome non-ATPase regulatory subunit	LSC00689	CN810710
5E-75	GTP-binding protein SAR1b	LSC00462	CN810361
1E-74	Dihydropteridine reductase	LSC00186	CN809979
2E-68	Calpain	LSC00545	CN810480
4E-67	Electron transfer flavoprotein alpha-subunit	LSC00720	CN810761
2E-65	Heterogeneous nuclear ribonucleoprotein H	LSC00030	CN809750
4E-64	Tyrosine/tryptophan monooxygenase activation protein	LSC00032	CN809753
4E-63	Arginase	LSC00265	CN810090
4E-63	Vacuolar ATP synthase	LSC00334	CN810177
7E-61	CTP synthase	LSC00045	CN809767
2E-59	Methionine adenosyltransferase	LSC00887	CN810989
4E-58	Phosphoglycerate mutase	LSC00814	CN810901
2E-56	Glutaryl-Coenzyme A dehydrogenase	LSC00283	CN810109
3E-55	Zinc finger protein 265	LSC00167	CN809952
5E-52	RER1 protein (Retention of ER protein 1)	LSC00373	CN810222
8E-52	Ubiquinol-Cytochrome c reductase	LSC00379	CN810236
3E-49	PBX1 (extradenticle)	LSC00198	CN809997
7E-48	Sorting nexin 17	LSC00492	CN810406
1E-43	Triose phosphate isomerase	LSC00097	CN809868
4E-41	Spectrin beta chain	LSC00205	CN810004
2E-40	Eukaryotic translation initiation factor 3	LSC00064	CN809793
1E-39	Annexin	LSC00885	CN811006
6E-35	Eyes absent 1	LSC00760	CN810823
1E-34	Laccase 1	LSC000184	CN809977
1E-34	Acyl carrier protein	LSC00346	CN810190
3E-31	Spectrin alpha chain	LSC00427	CN810307
4E-29	60S ribosomal protein L24	LSC00153	CN809905
8E-27	DNA repair protein RAD52 homologue	LSC00615	CN810583
4E-13	Amyloid protein	LSC00321	CN810160

[1] zeldia.cap.ed.ac.uk/mollusca.html.
[2] dbEST reference for first ocurence of sequence.

haemolymph proteins and cytokine-like components also contribute to resistance (Bayne *et al*. 2001). For example, upon exposure to infection by trematodes, *B. glabrata* snails increase production of fibrinogen-related proteins (FREPs) in the haemolymph, which recognize and precipitate trematode antigens (Adema *et al*. 1997, 1999; Léonard *et al*. 2001; Zhang *et al*. 2004). We did not identify any expressed FREPs in *L. stagnalis*, which is perhaps not surprising since the snails had not been exposed to a parasite, and the libraries were CNS-derived.

Differential display experiments have also shown that reverse transcriptases are abundantly expressed in infected *B. glabrata*, but only at moderate levels in

Table 4. Genes in the *Lymnaea stagnalis* EST data set that derive from the mitochondrial genome

Gene	No. of ESTs	Cluster No.[1]	GenBank[2]
Cytochrome c oxidase subunit I	11	LSC00003	CN809745
16S rRNA	118	LSC00004	CN809741
NADH dehydrogenase subunit IV	3	LSC00008	CN809861
Cytochrome c oxidase subunit II	5	LSC00348	CN810811
ATP synthase subunit VI	1	LSC00109	CN810526
NADH dehydrogenase subunit II	2	LSC00120	CN809884
12S rRNA	1	LSC00890	CN811015

[1] zeldia.cap.ed.ac.uk/mollusca.html.
[2] dbEST reference for first occurence of sequence.

uninfected snails (Raghavan *et al*. 2003). The reverse transcriptases probably derive from endogenous retrovirus activity, suggesting that infected *B. glabrata* may be a compromised host in which pathogens are more free to replicate. Reverse transcriptase sequences were also rare in the *L. stagnalis* EST dataset (CN809742, CN809820, CN810394, CN811017 only), but their presence is still intriguing.

More significantly, we isolated heat shock protein (HSP70; clusters LSC00649, LSC00099; GenBank AF025477), which has been shown to be upregulated in schistosome-resistant *B. glabrata* (Jones *et al*. 2001; Lockyer *et al*. 2004). On a similar note, we also isolated a gene with globin domains from *L. stagnalis* (LSC00360), an orthologue of which is again differentially expressed in *B. glabrata* (Lockyer *et al*. 2004). Similarly, several neuropeptides that were also isolated in *L. stagnalis*, including pedal peptide, show divergent expression during parasitosis of *B. glabrata*, perhaps as a result of the parasite modulating snail behaviour (Hoek *et al*. 1997).

Recently, Nowak *et al*. (2004) used suppression subtractive hybridization to enrich for transcripts that are expressed in a resistant strain of *B. glabrata*. Eighty-eight unique ESTs were isolated, and further screening showed that 22 of these were significantly up-regulated in exposed infected snails. The majority of the ESTs were novel. Comparing these putative resistance genes to the *L. stagnalis* dataset identified 7 putative orthologues, 2 of which were identified by Nowak *et al*. (2004) as differentially expressed. One gene, cytochrome c oxidase subunit VIb (CN809827, cluster LSC00086 in *L. stagnalis*; CD760681 in *B. glabrata*), is also associated with an RFLP marker that distinguishes resistant from susceptible strains of snails (Knight *et al*. 1998). The other is a gene of unknown function (CN810899, cluster LSC00764 in *L. stagnalis*; CD760608 in *B. glabrata*), with no homologues in any other organisms.

Several other putative immune defence genes were also isolated, including an orthologue of achacin from *Achatina fulica* (Obara *et al*. 1992), and ink toxin or aplysianin from the sea-hare (*Aplysia punctata*; Butzke *et al*. 2004), both of which are antibacterial amino oxidases (Tossi & Sandri, 2002). Multiple copies of a *B. glabrata* achacin/aplysianin orthologue are also present in dbEST. Finally, Knight *et al*. (1999) used genetic crosses to identify two RAPD markers (both repetitive sequences within the genome) that co-segregate with resistance to schistosomes. Similar repetitive sequences are not present in our EST data set.

Neuropeptides

Many of the most abundant sequences from our *L. stagnalis* CNS cDNA library corresponded to previously identified *L. stagnalis* sequences, though some, such as ovulation prohormone precursor and APGWamide, are known from their protein sequence only (Ebberink *et al*. 1985; Smit *et al*. 1992). Moreover, at least 8 of the most abundant ESTs (Table 1) corresponded to known neuropeptides (e.g. APGWamide, FMRFamide, Smit *et al*. 1992, 1993; preproLYCP, Kellett *et al*. 1994). One *L. stagnalis* cluster (CN810391, CN810195; LSC00324) had 100% amino acid identity (63% DNA) with the achatin neuropeptide, a gene previously isolated in *Achatina fulica* (Giant African land snail; Satake *et al*. 1999) and *Helix lucorum* (garden snail). Another *L. stagnalis* cluster (CN809782; LSC00018) had 100% identity (over 46 amino acids) with APGWamide, first isolated in *Aplysia californica* (U85585). Additional neuropeptide genes were also identified in the lower copy number clusters, bringing the known total to more than 30.

Few of the previously identified *L. stagnalis* neuropeptides had significant similarity to *B. glabrata* sequences, probably because no CNS ESTs have been generated from the latter snail, and the expression of many genes will be tissue-specific. Instead, many of the *L. stagnalis* and *B. glabrata* joint hits were conserved ribosomal or structural proteins, which are presumably expressed in most tissues. As studies on *B. glabrata* have thus far focused on haemocytes (e.g. Raghavan *et al*. 2003) and other tissues including the ovotestis and haemopoetic organ, then this emphasizes that a range of other tissue-specific libraries should be sampled to fully elucidate the transcriptome.

Genes of unknown function

Of particular interest for future research are the abundant ESTs that have no sequence similarity to other genes (~52%), yet which must still have an

important function. For example, one EST cluster from *L. stagnalis*, proved to contain a remarkable diversity of different, but closely related sequences (10 sequences were derived). It will be interesting to see whether the same gene family exists in *B. glabrata* (which may require a CNS EST library, or else a direct search) and more distantly related snails and molluscs.

Lophotrochozoan genes

From an evolutionary perspective snails are part of a large clade, the Lophotrochozoa, that includes not only their trematode platyhelminth parasites, but also annelids (e.g. earthworms), bryozoans, and rotifers, amongst other phyla (Winnepenninckx *et al.* 1995; Peterson & Eernisse, 2001). The group in general has been under-represented by recent genome sequencing efforts, except for the schistosome genome project and an earthworm EST project (www.earthworms.org; Sturzenbaum *et al.* 2003). The situation is rapidly changing with the recent addition of a large number of oyster *Crassostrea* sp. ESTs (>5000; Jenny *et al.* 2002; Gueguen *et al.* 2003) and the drive to sequence the *B. glabrata* genome (http://www.genome.gov/11007951). Thus, it is not surprising that several genes were identified in the *L. stagnalis* EST dataset that have not been previously described from any lophotrochozoan species. As about 48% of the clusters hit another protein in the non-redundant databases, and only 24% were significantly similar to any lophotrochozoan sequences, we estimate that around 190 genes were isolated for the first time in the Lophotrochozoa, several of which are especially interesting from an evolutionary perspective.

The homeobox protein PBX1 (known in *Drosophilamelanogaster* as extradenticle) is a transcription factor involved in development that has been described previously in both the Ecdysozoa and Deuterostomia (Monica *et al.* 1991; Rauskolb, Peifer& Wieschaus, 1993). The cDNA clone corresponding to the EST was recovered from stocks and sequenced using internal primers, yielding a complete open reading frame and ~230 bp of the 5′ untranslated region (CN809997). Interestingly, though there are many more sequences from the earthworm (*Lumbricus rubellus*; ~12 000) and *Schistosoma mansoni* (~156 000) in GenBank, the earthworm PBX1 has not been isolated and there is only one putative PBX1 EST from *S. mansoni*. Similarly, an alpha coat protein (involved with the traffic of proteins through the secretory pathway), a dynein (microtubule-based motor protein), a ruvB-like protein (branch migration of Holliday junctions), and finally a DEAD-box peptide were also novel lophotrochozoan isolates. Another interesting cluster was a putative glycosylhydrolase family 9 gene (GHF9). GHF9 endoglucanases degrade cellulose,

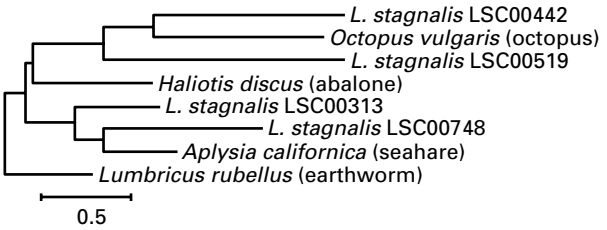

Fig. 5. Neighbour-joining DNA phylogeny of paralogous *beta-tubulin* genes (665 bp). The 4 *beta tubulin* paralogues from *L. stagnalis* may be the result of ancient duplication.

yet few orthologues have previously been isolated from the Metazoa, giving rise to the suggestion that the gene was gained by horizontal gene transfer from bacteria, though other explanations are possible (Lo, Watanabe & Sugimura, 2003). This clone was also recovered and sequenced. Initial analyses indicate that it is the orthologue of a previously identified abalone (*Haliotisdiscus*) gene, lending support to the hypothesis that these GHF9 genes were anciently present in animals (Lo*et al.* 2003).

Genes for phylogenies

Some of the existing *L. stagnalis* DNA sequences were produced for phylogenetic reconstruction, in particular mitochondrial DNA sequences. Previously, these analyses have been limited to 16S rRNA and cytochrome oxidase subunit I genes, because of the lack of conserved primers available for molluscs (Remigio & Blair, 1997; Remigio & Hebert, 2003). The EST data set contained 5 additional mitochondrial genes: NADH dehydrogenase subunits II and IV, cytochrome oxidase subunit II, ATP synthase subunit VI, and 12S rRNA, which can now be used, in conjunction with complete and partial mitochondrial genomes from other taxa, to develop further conserved primer sets. Sequencing and characterization of the complete *L. stagnalis* mitochondrial genome should also be straightforward. The *B. glabrata* mitochondrial genome has recently been sequenced (Dejong, Emery & Adema, 2004), and adding a *L. stagnalis* genome to the dataset could improve our understanding of the evolution of snails and their relationships with parasites.

For similar reasons, several of the ESTs may be useful for phylogenetic reconstruction at a variety of systematic levels. Highly conserved genes may be useful for recovery of branching orders of deep divergences. For example, a catenated dataset of beta-tubulin, alpha-tubulin, elongation factor 1 alpha (EF1alpha) and actin was used to attempt to resolve the higher order systematics of Eukaryota (Baldauf *et al.* 2000). Beta-tubulin has also been used to investigate microsporidian/fungal phylogeny (Keeling, Luker & Palmer, 2000) and the relationships between 'jakobid' flagellates (Edgcomb *et al.*

2001). However, 4 distinct beta-tubulin paralogues were recovered in the *L. stagnalis* ESTs. Phylogenetic analysis of these genes with other mollusc beta-tubulins revealed that at least 1 of the gene duplications that generated these 4 paralogues in *L. stagnalis* may be ancient (Fig. 5). This argues for caution in the use of this gene in phylogenetic analysis. In contrast, EF1alpha has almost always been found to be a single copy gene and this study, the earthworm EST project, and the schistosome genome project have only recovered a single copy of EF1alpha. Phylogenetic analysis using EF1alpha (LSC00016, LSC00124) recovers the expected tree (not shown).

Mapping strategies

Finally, these ESTs may be useful in illuminating some of the many other fascinating aspects of snail biology. *L. stagnalis* is a model for the evolution of body asymmetry because populations are sometimes polymorphic for shell and body asymmetry (sinistral and dextral coiling; Hosoiri *et al.* 2003; Shibazaki, Shimizu & Kuroda, 2004). This coiling asymmetry has been shown to be genetically determined, and it will be interesting to isolate the gene(s) responsible and compare them to orthologues in humans and nematodes. The ESTs provide a source of marker loci that could be developed into PCR fragment size, restriction fragment length polymorphism or single nucleotide polymorphism markers to facilitate genetic mapping (e.g. Choi *et al.* 2004; Komulainen *et al.* 2003).

These ESTs will also provide additional molecular markers for analysis of the function of the snail CNS, the molecular components of behaviour and response to parasitism. In particular, the diversity of neuropeptides discovered by EST sequencing suggests that neurohormonal control may be more complex than previously modelled. Comparative analysis with the *Biomphalaria*/*Schistosoma* pair will help to identify both conserved and divergent aspects of the host-parasite relationship.

This work was supported by funding from the Royal Society (A.D.), NERC and Wellcome (M.B.) for which the authors are very grateful. Thanks to Jen Daub, Claire Whitton, Katelyn Fenn, James Wasmuth, Alasdair Anthony, and Ralf Schmid for technical assistance. Ronald van Kesteren and Hisayo Sadamoto kindly supplied the cDNA libraries.

REFERENCES

ADEMA, C. M., HERTEL, L. A., MILLER, R. D. & LOKER, E. S. (1997). A family of fibrinogen-related proteins that precipitate parasite-derived molecules is produced by an invertebrate after infection. *Proceedings of the National Academy of Sciences, USA* **94**, 8691–8696.

ADEMA, C. M., HERTEL, L. A. & LOKER, E. S. (1999). Evidence from two planorbid snails of a complex and dedicated response to digenean (echinostome) infection. *Parasitology* **119**, 395–404.

ADEMA, C. M. (2002). Comparative study of cytoplasmic actin DNA sequences from six species of Planorbidae (Gastropoda: Basommatophora). *Journal of Molluscan Studies* **68**, 17–23.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.

ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402.

BALDAUF, S. L., ROGER, A. J., WENK-SIEFERT, I. & DOOLITTLE, W. F. (2000). A. kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**, 972–977.

BARGUES, M. D., HORAK, P., PATZNER, R. A., POINTIER, J. P., JACKIEWICZ, M., MEIER-BROOK, C. & MAS-COMA, S. (2003). Insights into the relationships of palearctic and nearctic lymnaeids (Mollusca: Gastropoda) by rDNA ITS-2 sequencing and phylogeny of Stagnicoline intermediate host species of *Fasciola hepatica*. *Parasite-Journal de La Societé Francaise de Parasitologie* **10**, 243–255.

BAYNE, C. J., HAHN, U. K. & BENDER, R. C. (2001). Mechanisms of molluscan host resistance and of parasite strategies for survival. *Parasitology* **123**, S159–S167.

BENDTSEN, J. D., NIELSEN, H., VON HEIJNE, G. & BRUNAK, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* **340**, 783–795.

BRUNET, J. F., SHAPIRO, E., FOSTER, S. A., KANDEL, E. R. & IINO, Y. (1991). Identification of a peptide specific for *Aplysia* sensory neurons by PCR-based differential screening. *Science* **252**, 856–859.

BUTZKE, D., MACHUY, N., THIEDE, B., HURWITZ, R., GOEDERT, S. & RUDEL, T. (2004). Hydrogen peroxide produced by *Aplysia* ink toxin kills tumor cells independent of apoptosis via peroxiredoxin I sensitive pathways. *Cell Death and Differentiation* **11**, 608–617.

CHASE, R. (2002). *Behavior and its Neural Control in Gastropod Molluscs*. Oxford University Press, Oxford.

CHOI, H. K., KIM, D., UHM, T., LIMPENS, E., LIM, H., MUN, J. H., KALO, P., PENMETSA, R. V., SERES, A., KULIKOVA, O., ROE, B. A., BISSELING, T., KISS, G. B. & COOK, D. R. (2004). A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. *Genetics* **166**, 1463–1502.

VON DARL, M., HARRISON, P. M. & BOTTKE, W. (1994). cDNA cloning and deduced amino acid sequence of two ferritins: soma ferritin and yolk ferritin, from the snail *Lymnaea stagnalis*. *European Journal of Biochemistry* **222**, 353–366.

DAVISON, A. (2002). Land snails as a model to understand the role of history and selection in the origins of biodiversity. *Population Ecology* **44**, 129–136.

DAUB, J., LOUKAS, A., PRITCHARD, D. I. & BLAXTER, M. (2000). A survey of genes expressed in adults of the human hookworm, *Necator americanus*. *Parasitology* **120**, 171–184.

DEJONG, R. J., EMERY, A. M. & ADEMA, C. M. (2004). The mitochondrial genome of *Biomphalaria glabrata* (Gastropoda, Basommatophora), intermediate host of *Schistosoma mansoni Journal of Parasitology* (in the Press).

DEWILDE, S., WINNEPENNINCKX, B., ARNDT, M. H. L., NASCIMENTO, D. G., SANTORO, M. M., KNIGHT, M., MILLER, A. N., KERLAVAGE, A. R., GEOGHAGEN, N., VAN MARCK, E., LIU, L. X., WEBER, R. E. & MOENS, L. (1998). Characterization of the myoglobin and its coding gene of the mollusc *Biomphalaria glabrata*. *Journal of Biological Chemistry* **273**, 13583–13592.

DUCKERT, P., BRUNAK, S. & BLOM, N. (2004). Prediction of proprotein convertase cleavage sites. *Protein Engineering, Design and Selection* **17**, 107–112.

EBBERINK, R. H. M., VAN LOENHOUT, H., GERAERTS, W. P. M. & JOOSSE, J. (1985). Purification and amino acid sequence of the ovulation neurohormone of *Lymnaea stagnalis*. *Proceedings of the National Academy of Sciences, USA* **82**, 7767–7771.

EDGCOMB, V. P., ROGER, A. J., SIMPSON, A. G. B., KYSELA, D. T. & SOGIN, M. L. (2001). Evolutionary relationships among ''jakobid'' flagellates as indicated by alpha- and beta-tubulin phylogenies. *Molecular Biology and Evolution* **18**, 514–522.

ESTEBAN, J. G., BARGUES, M. D. & MAS-COMA, S. (1998). Geographical distribution, diagnosis and treatment of human fascioliasis: a review. *Research and Reviews in Parasitology* **58**, 13–42.

FRANTSEVICH, L., KORNIUSHIN, A., PANKOV, I., ERMAKOV, A. & ZAKHARCHUK, T. (1996). Application of molluscs for radioecological monitoring of the Chernobyl outburst. *Environmental Pollution* **94**, 91–100.

GUEGUEN, Y., CADORET, J. P., FLAMENT, D., BARREAU-ROUMIGUIERE, C., GIRARDOT, A. L., GARNIER, J., HOAREAU, A., BACHERE, E. & ESCOUBAS, J. M. (2003). Immune gene discovery by expressed sequence tags generated from hemocytes of the bacteria-challenged oyster, *Crassostrea gigas*. *Gene* **303**, 139–145.

HENNIG, S., GROTH, D. & LEHRACH, H. (2003). Automated gene ontology annotation for anonymous sequence data. *Nucleic Acids Research* **31**, 3712–3715.

HOEK, R. M., VAN KESTEREN, R. E., SMIT, A. B., DE JONG-BRINK, M. & GERAERTS, W. P. (1997). Altered gene expression in the host brain caused by a trematode parasite: neuropeptide genes are preferentially affected during parasitosis. *Proceedings of the National Academy of Sciences, USA* **94**, 14072–14076.

HOSOIRI, Y., HARADA, Y. & KURODA, R. (2003). Construction of a backcross progeny collection of dextral and sinistral individuals of a freshwater gastropod, *Lymnaea stagnalis*. *Development Genes and Evolution* **213**, 193–198.

HURTREZ-BOUSSES, S., MEUNIER, C., DURAND, P. & RENAUD, F. (2001). Dynamics of host-parasite interactions: the example of population biology of the liver fluke (*Fasciola hepatica*) *Microbes and Infection* **3**, 841–849.

JENNY, M. J., RINGWOOD, A. H., LACY, E. R., LEWITUS, A. J., KEMPTON, J. W., GROSS, P. S., WARR, G. W. & CHAPMAN, R. W. (2002). Potential indicators of stress response identified by expressed sequence tag analysis of hemocytes and embryos from the American oyster, *Crassostrea virginica*. *Marine Biotechnology* **4**, 81–93.

JONES, C. S., LOCKYER, A. E., ROLLINSON, D. & NOBLE, L. R. (2001). Molecular approaches in the study of *Biomphalaria glabrata – Schistosoma mansoni* interactions: linkage analysis and gene expression profiling. *Parasitology* **123**, S181–S196.

KEELING, P. J., LUKER, M. A. & PALMER, J. D. (2000). Evidence from beta-tubulin phylogeny that microsporidia evolved from within the fungi. *Molecular Biology and Evolution* **17**, 23–31.

KELLETT, E., SAUNDERS, S. E., LI, K. W., STADDON, J. W., BENJAMIN, P. R. & BURKE, J. F. (1994). Genomic organization of the FMRFamide gene in *Lymnaea*: multiple exons encoding novel neuropeptides. *Journal of Neuroscience* **14**, 6564–6570.

KENYON, F., WELSH, M., PARKINSON, J., WHITTON, C., BLAXTER, M. L. & KNOX, D. P. (2003). Expressed sequence tag survey of gene expression in the scab mite *Psoroptes ovis* – allergens, proteases and free-radical scavengers. *Parasitology* **126**, 451–460.

KNIGHT, M., MILLER, A. N., GEOGHAGEN, N. S. M., LEWIS, F. A. & KERLAVAGE, A. R. (1998). Expressed sequence tags (ESTs) of *Biomphalaria glabrata*, an intermediate snail host of *Schistosoma mansoni*: use in the identification of RFLP markers. *Malacologia* **39**, 175–182.

KNIGHT, M., MILLER, A. N., PATTERSON, C. N., ROWE, C. G., MICHAELS, G., CARR, D., RICHARDS, C. S. & LEWIS, F. A. (1999). The identification of markers segregating with resistance to *Schistosoma mansoni* infection in the snail *Biomphalaria glabrata*. *Proceedings of the National Academy of Sciences, USA* **96**, 1510–1515.

KNIGHT, M., ONGELE, E. & LEWIS, F. A. (2000). Molecular studies of *Biomphalaria glabrata*, an intermediate host of *Schistosoma mansoni*. *International Journal for Parasitology* **30**, 535–541.

KOMULAINEN, P., BROWN, G. R., MIKKONEN, M., KARHU, A., GARCIA-GIL, M. R., O'MALLEY, D., LEE, B., NEALE, D. B. & SAVOLAINEN, O. (2003). Comparing EST-based genetic maps between *Pinus sylvestris* and *Pinus taeda*. *Theoretical And Applied Genetics* **107**, 667–678.

LÉONARD, P. M., ADEMA, C. M., ZHANG, S. & LOKER, E. S. (2001). Structure of two FREP genes that combine IgSF and fibrinogen domains, with comments on diversity of the FREP gene family in the snail *Biomphalaria glabrata*. *Gene* **269**, 155–165.

LO, N., WATANABE, H. & SUGIMURA, M. (2003). Evidence for the presence of a cellulase gene in the last common ancestor of bilaterian animals. *Proceedings of the Royal Society, Series B* **270**, S69–S72.

LOCKHART, S. T., LEVITAN, I. B. & PIKIELNY, C. W. (1996). Ag, a novel protein secreted from *Aplysia* glia. *Journal of Neurobiology* **29**, 35–48.

LOCKYER, A. E., NOBLE, L. R., ROLLINSON, D. & JONES, C. S. (2004). *Schistosoma mansoni*: resistant specific infection-induced gene expression in *Biomphalaria glabrata* identified by fluorescent-based differential display. *Experimental Parasitology* **107**, 97–104.

LUKOWIAK, K., SANGHA, S., MCCOMB, C., VARSHNEY, N., ROSENEGGER, D., SADAMOTO, H. & SCHEIBENSTOCK, A. (2003). Associative learning and memory in *Lymnaea stagnalis*: how well do they remember? *Journal of Experimental Biology* **206**, 2097–2103.

MAS-COMA, S., FUNATSU, I. R. & BARGUES, M. D. (2001). *Fasciola hepatica* and lymnaeid snails occurring at very high altitude in South America. *Parasitology* **123**, S115–S127.

MEUNIER, C., HURTREZ-BOUSSES, S., JABBOUR-ZAHAB, R., DURAND, P., RONDELAUD, D. & RENAUD, F. (2004). Field

and experimental evidence of preferential selfing in the freshwater mollusc *Lymnaea truncatula* (Gastropoda, Pulmonata). *Heredity* **92**, 316–322.

MILLER, A. N., RAGHAVAN, N., FITZGERALD, P. C., LEWIS, F. A. & KNIGHT, M. (2001). Differential gene expression in haemocytes of the snail *Biomphalaria glabrata*: effects of *Schistosoma mansoni* infection. *International Journal for Parasitology* **31**, 687–696.

MOCCIA, R., CHEN, D., LYLES, V., KAPUYA, E., YAPING, E., KALACHIKOV, S., SPAHN, C. M., FRANK, J., KANDEL, E. R., BARAD, M. & MARTIN, K. C. (2003). An unbiased cDNA library prepared from isolated *Aplysia* sensory neuron processes is enriched for cytoskeletal and translational mRNAs. *Journal of Neuroscience* **23**, 9409–9417.

MONICA, K., GALILI, N., NOURSE, J., SALTMAN, D. & CLEARY, M. L. (1991). PBX2 and PBX3, new homeobox genes with extensive homology to the human proto-oncogene PBX1. *Molecular and Cellular Biology* **11**, 6149–6157.

NOTO, T. & ENDOH, H. (2004). A ''chimera'' theory on the origin of dicyemid mesozoans: evolution driven by frequent lateral gene transfer from host to parasite. *BioSystems* **73**, 73–83.

NOWAK, T. S., WOODARDS, A. C., JUNG, Y., ADEMA, C. M. & LOKER, E. S. (2004). Identification of transcripts generated during the response of resistant *Biomphalaria glabrata* to *Schistosoma mansoni* infection using suppression subtractive hybridization (SSH). *Journal of Parasitology* (in the Press).

OBARA, K., OTSUKA-FUCHINO, H., SATTAYASAI, N., NONOMURA, Y., TSUCHIYA, T. & TAMIYA, T. (1992). Molecular cloning of the antibacterial protein of the giant African snail, *Achatina fulica* Ferussac. *European Journal of Biochemistry* **209**, 1–6.

PARKINSON, J., ANTHONY, A., WASMUTH, J., SCHMID, R., HEDLEY, A. & BLAXTER, M. (2004). PartiGene – constructing partial genomes. *Bioinformatics* **20**, 1398–1404.

PETERSON, K. J. & EERNISSE, D. J. (2001). Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA sequences. *Evolution and Development* **3**, 170–205.

PIEDRAFITA, D., RAADSMA, H. W., PROWSE, R. & SPITHILL, T. W. (2004). Immunology of the host–parasite relationship in fasciolosis (*Fasciola hepatica* and *Fasciola gigantica*). *Canadian Journal of Zoology* **82**, 233–250.

RAGHAVAN, N., MILLER, A. N., GARDNER, M., FITZGERALD, P. C., KERLAVAGE, A. R., JOHNSTON, D. A., LEWIS, F. A. & KNIGHT, M. (2003). Comparative gene analysis of *Biomphalaria glabrata* hemocytes pre- and post-exposure to miracidia of *Schistosoma mansoni*. *Molecular and Biochemical Parasitology* **126**, 181–191.

RAUSKOLB, C., PEIFER, M. & WIESCHAUS, E. (1993). Extradenticle, a regulator of homeotic gene activity, is a homolog of the homeobox-containing human proto-oncogene pbx1. *Cell* **74**, 1101–1112.

REMIGIO, E. A. & BLAIR, D. (1997). Molecular systematics of the freshwater snail family Lymnaeidae (Pulmonata: Basommatophora) utilising mitochondrial ribosomal DNA sequences. *Journal of Molluscan Studies* **63**, 173–185.

REMIGIO, E. A. & HEBERT, P. D. (2003). Testing the utility of partial COI sequences for phylogenetic estimates of gastropod relationships. *Molecular Phylogenetics and Evolution* **29**, 641–647.

SADAMOTO, H., SATO, H., KOBAYASHI, S., MURAKAMI, J., AONUMA, H., ANDO, H., FUJITO, Y., HAMANO, K., AWAJI, M., LUKOWIAK, K., URANO, A. & ITO, E. (2004). CREB in the pond snail *Lymnaea stagnalis*: cloning, gene expression, and function in identifiable neurons of the central nervous system. *Journal of Neurobiology* **58**, 455–466.

SALANKI, J. (2000). Invertebrates in neurotoxicology. *Acta Biologica Hungarica* **51**, 287–307.

SATAKE, H., YASUDA-KAMATANI, Y., TAKUWA, K., NOMOTO, K., MINAKATA, H., NAGAHAMA, T., NAKABAYASHI, K. & MATSUSHIMA, O. (1999). Characterization of a cDNA encoding a precursor polypeptide of a D-amino acid-containing peptide, achatin-I and localized expression of the achatin-I and fulicin genes. *European Journal of Biochemistry* **261**, 130–136.

SHIBAZAKI, Y., SHIMIZU, M. & KURODA, R. (2004). Body handedness is directed by genetically determined cytoskeletal dynamics in the early embryo. *Current Biology* **14**, 1462–167.

SMIT, A. B., JIMENEZ, C. R., DIRKS, R. W., CROLL, R. P. & GERAERTS, W. P. (1992). Characterization of a cDNA clone encoding multiple copies of the neuropeptide APGWamide in the mollusk *Lymnaea stagnalis*. *Journal of Neuroscience* **12**, 1709–1715.

SMIT, A. B., HOEK, R. M. & GERAERTS, W. P. (1993). The isolation of a cDNA encoding a neuropeptide prohormone from the light yellow cells of *Lymnaea stagnalis*. *Cellular and Molecular Neurobiology* **13**, 263–270.

SMIT, A. B., SPIJKER, S., VAN MINNEN, J., BURKE, J. F., DE WINTER, F., VAN ELK, R. & GERAERTS, W. P. (1996). Expression and characterization of molluscan insulin-related peptide VII from the mollusc *Lymnaea stagnalis*. *Neuroscience* **70**, 589–596.

SMIT, A. B., SYED, N. I., SCHAAP, D., VAN MINNEN, J., KLUMPERMAN, J., KITS, K. S., LODDER, H., VAN DER SCHORS, R. C., VAN ELK, R., SORGEDRAGER, B., BREJC, K., SIXMA, T. K. & GERAERTS, W. P. M. (2001). A glia-derived acetylcholine-binding protein that modulates synaptic transmission. *Nature, London* **411**, 261–268.

SORENSEN, R. E. & MINCHELLA, D. J. (2001). Snail-trematode life history interactions: past trends and future directions. *Parasitology* **123**, S3–S18.

STURZENBAUM, S. R., PARKINSON, J., BLAXTER, M., MORGAN, A. J., KILLE, P. & GEORGIEV, O. (2003). The earthworm Expressed Sequence Tag project. *Pedobiologia* **47**, 447–451.

TEUNISSEN, Y., GERAERTS, W. P., VAN HEERIKHUIZEN, H., PLANTA, R. J. & JOOSSE, J. (1992). Molecular cloning of a cDNA encoding a member of a novel cytochrome P450 family in the mollusc *Lymnaea stagnalis*. *Journal of Biochemistry* **112**, 249–252.

TODA, H., YAZAWA, M., KONDO, K., HONMA, T., NARITA, K. & YAGI, K. (1981). Amino acid sequence of calmodulin from scallop (*Patinopecten*) adductor muscle. *Journal of Biochemistry* **90**, 1493–1505.

TOSSI, A. & SANDRI, S. (2002). Molecular diversity in gene-encoded, cationic antimicrobial polypeptides. *Current Pharmaceutical Design* **8**, 743–761.

VIGNOLES, P., DREYFUSS, G. & RONDELAUD, D. (2002). Larval development of *Fasciola hepatica* in experimental infections: variations with populations of *Lymnaea truncatula*. *Journal of Helminthology* **76**, 179–183.

WINNEPENNINCKX, B., BACKELJAU, T., MACKEY, L. Y., BROOKS, J. M., DEWACHTER, R., KUMAR, S. & GAREY, J. R. (1995). 18S ribosomal-RNA data indicate that Aschelminthes are polyphyletic in origin and consist of at least three distinct clades. *Molecular Biology and Evolution* **12**, 1132–1137.

ZHANG, S. M., ADEMA, C. M., KEPLER, T. B. & LOKER, E. S. (2004). Diversification of Ig superfamily genes in an invertebrate. *Science* **305**, 251–254.