

## RESEARCH ARTICLE

## Open Access

# Molecular insights into land snail neuropeptides through transcriptome and comparative gene analysis

Kevin J Adamson<sup>1</sup>, Tianfang Wang<sup>1</sup>, Min Zhao<sup>1</sup>, Francesca Bell<sup>1</sup>, Anna V Kuballa<sup>1</sup>, Kenneth B Storey<sup>2</sup> and Scott F Cummins<sup>1\*</sup>

## Abstract

**Background:** Snails belong to the molluscan class Gastropoda, which inhabit land, freshwater and marine environments. Several land snail species, including *Theba pisana*, are crop pests of major concern, causing extensive damage to agriculture and horticulture. A deeper understanding of their molecular biology is necessary in order to develop methods to manipulate land snail populations.

**Results:** The present study used *in silico* gene data mining of *T. pisana* tissue transcriptomes to predict 24,920 central nervous system (CNS) proteins, 37,661 foot muscle proteins and 40,766 hepatopancreas proteins, which together have 5,236 unique protein functional domains. Neuropeptides, metabolic enzymes and epiphragmin genes dominated expression within the CNS, hepatopancreas and muscle, respectively. Further investigation of the CNS transcriptome demonstrated that it might contain as many as 5,504 genes that encode for proteins destined for extracellular secretion. Neuropeptides form an important class of cell-cell messengers that control or influence various complex metabolic events. A total of 35 full-length neuropeptide genes were abundantly expressed within *T. pisana* CNS, encoding precursors that release molluscan-type bioactive neuropeptide products. These included achatin, allatotropin, conopressin, elevenin, FMRFamide, LFRFamide, LRFNVamide, myomodulins, neurokinin Y, PKYMDT, PXFamide, sCAPamides and several insulin-like peptides. Liquid chromatography-mass spectrometry of neural ganglia confirmed the presence of many of these neuropeptides.

**Conclusions:** Our results provide the most comprehensive picture of the molecular genes and proteins associated with land snail functioning, including the repertoire of neuropeptides that likely play significant roles in neuroendocrine signalling. This information has the potential to expedite the study of molluscan metabolism and potentially stimulate advances in the biological control of land snail pest species.

**Keywords:** Snail, *Theba pisana*, Neuropeptides, Central nervous system, Muscle, Hepatopancreas

## Background

Molluscs are the second largest animal phylum, comprising about 7% of living animals and occupying habitats ranging from high alpine regions to deep sea vents, with a diverse range of lifestyles including predatory, scavenging, herbivorous, detritivorous and filter-feeding [1]. The most abundant class of mollusc are the gastropods, which include the land snails that have evolved

independently to life on land. Such a transition has required adaptations towards water saving and breathing dry air.

The central nervous system (CNS), hepatopancreas (or digestive gland) and foot muscle are all key organs controlling or having some influence on a snail's metabolic rate, whereby a diverse array of cellular modifications is necessary to regulate and co-ordinate changes at the organismal level. Many of these modifications need to be global, and tightly controlled. Such wide ranging control is typically achieved by chemical signalling using neurohormones, many of which are peptides [2]. Neuropeptides,

\* Correspondence: [scummins@usc.edu.au](mailto:scummins@usc.edu.au)

<sup>1</sup>Genecology Research Centre, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Maroochydore DC, Queensland 4558, Australia

Full list of author information is available at the end of the article

which occur in all animals with a nervous system [3], have been widely studied since the early 1950's when oxytocin and vasopressin neuropeptides were first identified and characterised [4,5]. Many neuropeptides derive from larger inactive precursors which are proteolytically cleaved into a number of smaller bioactive neuropeptides, generally all the same, but occasionally with different biological actions [3,6].

In molluscs, most neuropeptide research has been on the aquatic gastropods, *Aplysia*, *Lymnaea* and *Lottia*. In the marine gastropod, *Aplysia*, several neuropeptides have been well described that are involved with reproduction and learning [7-9]. In *A. californica*, the best known neuropeptide is FMRF-amide, which appears to provide physiological control of gills [10], and has also been found in heart tissue [11], along with the small cardioactive peptides (SCP<sub>A</sub> and SCP<sub>B</sub>) which increase heart rate and the amplitude of the beat [11]. The freshwater snail *L. stagnalis* has also been a model for mollusc neuropeptide research where studies have discovered five genes coding for the neuropeptides APGWamide, neuropeptide Y, conopressin, molluscan insulin-related peptide, and pedal peptide that are involved in muscle contraction and modulation in males [12,13]. *In silico* genome and transcriptome database mining have proven effective for high-throughput annotation of the presence and expression of neuropeptides in *Lottia*, as well as in bivalve oysters [9,14]. Less is known about the neuropeptide repertoire of terrestrial pulmonate gastropods, although recently, mass spectrometry has been used to identify and determine changes in neuropeptide profiles in the brain and haemolymph of the snail *Helix pomatia* during activity versus hibernation [15]. In that study, 19 neuropeptides were identified as being more highly produced within the brain of active snails. Meanwhile, during hibernation, 11 neuropeptides were exclusively present [15].

In the current study, we investigated the CNS, hepatopancreas and foot muscle transcriptomes of the land snail *Theba pisana* through gene and peptide analysis. We found numerous neuropeptide precursors that show similarity with other known molluscan neuropeptide precursors, and also demonstrate the existence of numerous other genes that encode peptides destined for secretion. This represents the most extensive analysis of neuropeptide genes and their products in a terrestrial snail.

## Results

### De novo assembly and comparison of *Theba* CNS, foot and hepatopancreas

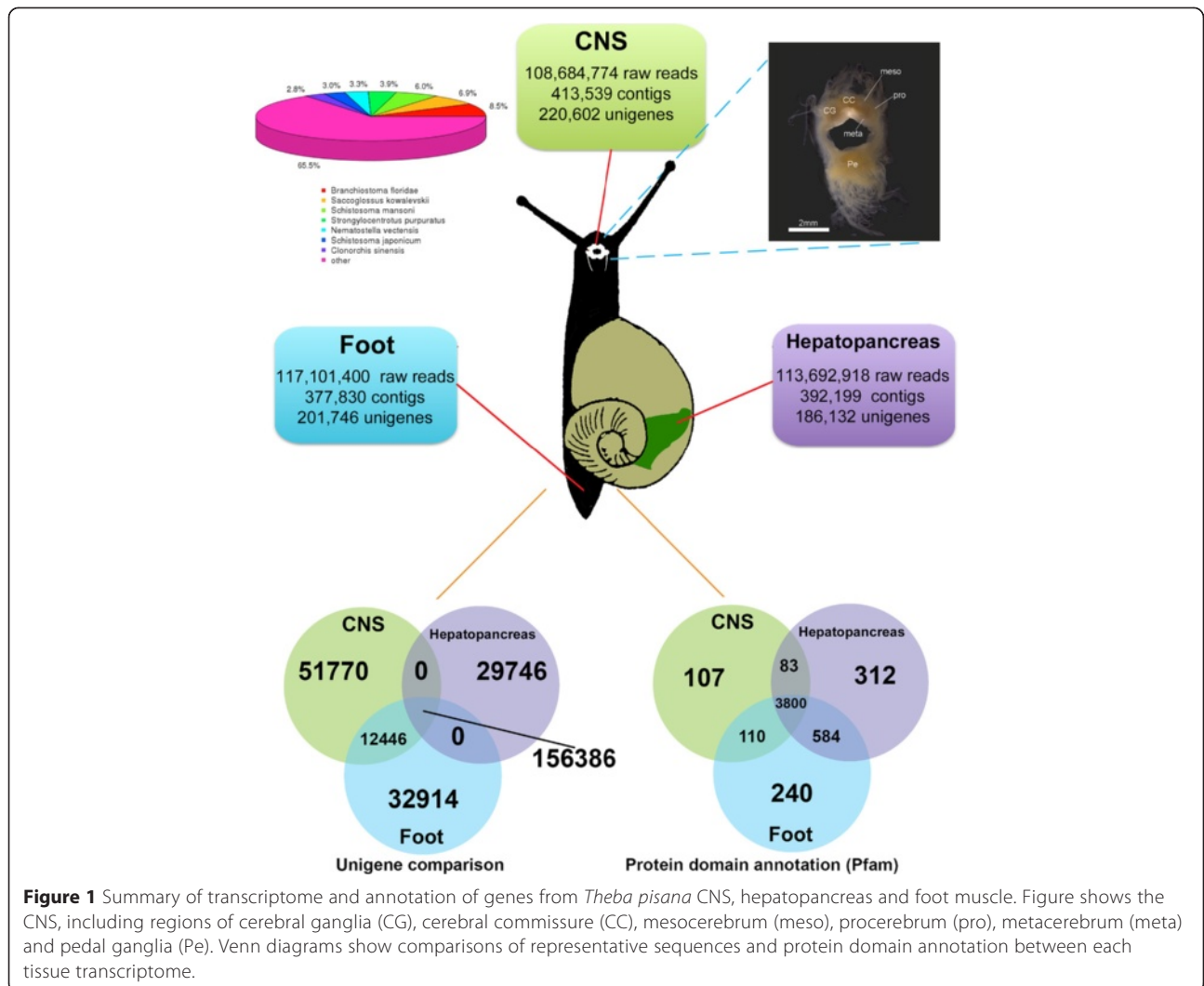
Transcript libraries derived from *T. pisana* CNS, foot and hepatopancreas tissues were sequenced using Illumina technologies and assembled. All sequence data was deposited in the NCBI Genbank under SRA file SRP056280. A summary of the number of high quality raw reads, contigs

and unigenes for each tissue is shown in Figure 1. The CNS, foot and hepatopancreas transcriptomes encoded 220,602, 201,746, and 186,132 unigenes, respectively. A unigene is typically interpreted as representing a single genomic locus; hence, these groups represent the first comprehensive non-redundant putative transcript database for *T. pisana*. Collectively, contigs and unigenes could be assembled into a total of 250,848 consensus sequences ranging in size from 200 to 26,246 bp; size distribution for the pooled tissue transcriptome is shown in Additional file 1: Figure S1. Approximately 156,386 representative sequences are shared within the three transcriptomes (Figure 1). Although CNS has more unique unigenes compared to foot or hepatopancreas, there was no substantial difference based on protein functional domain annotation (Figure 1). In total, there were 24,920 CNS proteins, 37,661 foot proteins and 40,766 hepatopancreas proteins annotated, with 5236 unique Pfam domains. Combining the unigene and Pfam comparison results, the data revealed that the three tissues have a moderate number of tissue-specific transcripts that encode proteins that play similar functions in cellular processes.

### Functional annotation

*T. pisana* sequences were annotated against protein databases (Nr, Nt, SwissProt, KEGG, GO and COG) using BLASTX (E-value < 0.00001). From the 250,848 consensus sequences, 69,799 (27.8%) had at least one hit. The sequence names and annotation information of all sequences are provided in Additional file 2: Table S1. The majority of transcripts had either a significant match with those from the pacific oyster (*Crassostrea gigas*) or did not match to any known genes; this result is most likely due to insufficient sequences being available in public databases from phylogenetically closely related species. The annotation rate in our study is comparable to those that have been reported in previous *de novo* transcriptome sequencing studies for molluscs [16,17].

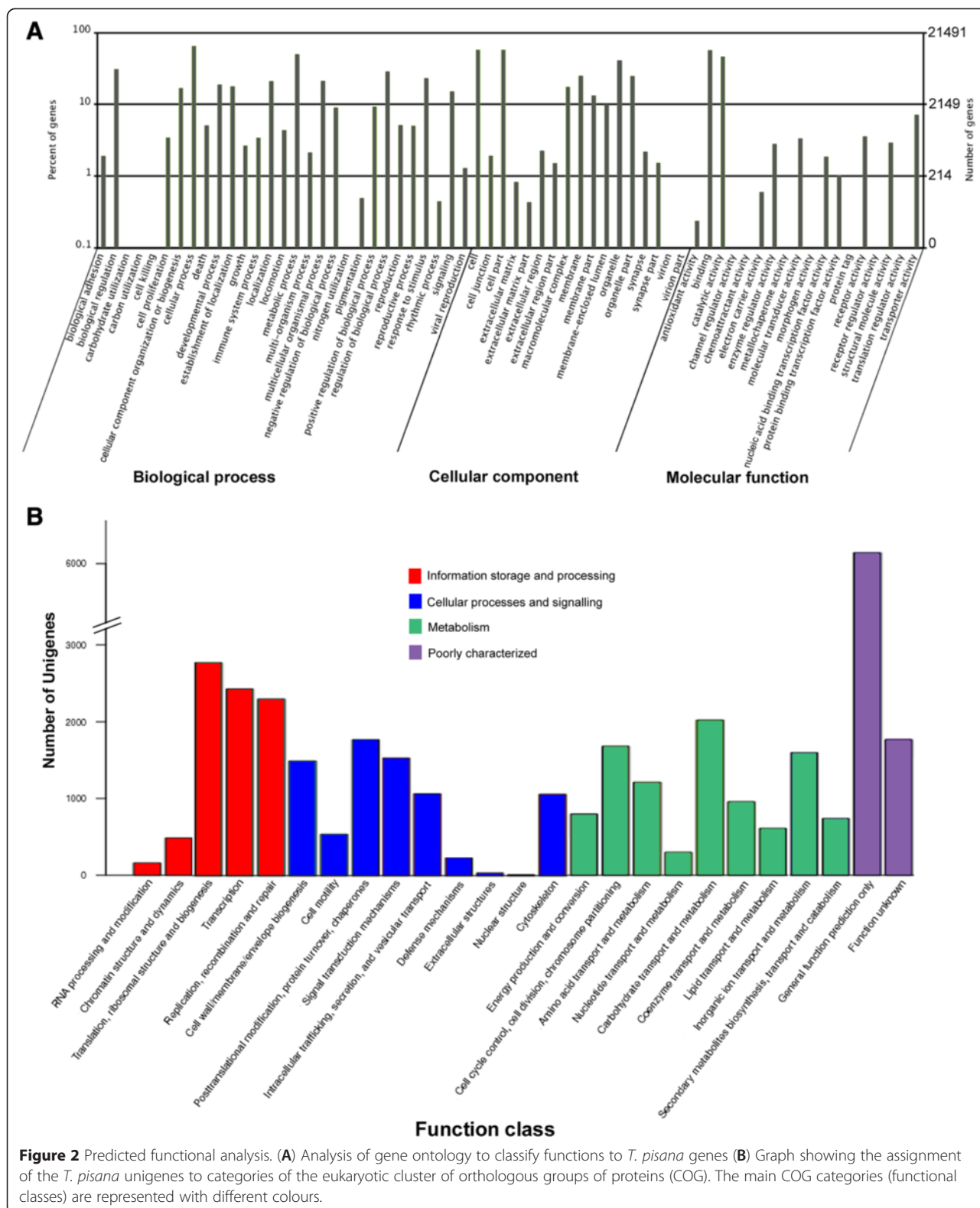
Gene ontology was performed to classify functions to *T. pisana* genes (Figure 2A). Of these, 77,715 transcripts were assigned to functional categories of 'Biological Process' (48.5%), 27,187 to 'Molecular Function' (34.5%) and 55,346 to 'Cellular Component' (17%). Functional annotation of all transcripts combined against the cluster of orthologous groups (COG) database is shown in Figure 2B. These were assigned to four primary COG classes: Information storage and processing (8156 transcripts), Cellular processes and signalling (7756 transcripts), and Metabolism (9932 transcripts) as well as Poorly characterized genes (7,911 transcripts). The COG functional classification demonstrates that the most abundant classification is "general function prediction only", followed by "translation, ribosomal structure and biogenesis" and "transcription".



Transcript abundance was determined based on fragments per kilobase of exon per million fragments mapped (FPKM; Additional file 1: Figure S2 and Additional file 2: Table S1). In the CNS, of the 50 most abundant transcripts (besides ribosomal), 22% encoded for neuropeptides such as insulin, neuropeptide Y, myomodulin and achatin, whereas unannotated transcripts comprised 40% of transcripts. In the hepatopancreas, many of the top 50 abundant transcripts (minus ribosomal) encoded for catabolic enzymes including cathepsin peptidase, serine peptidase, chitotriosidase-1, myosinase, lysozyme, while 20% were unannotated including the most abundant transcript (Unigene64357\_All). In the foot muscle, the epiphragmin-encoding transcripts were most prominent. Common transcripts that were of high abundance in all three transcriptomes included those encoding a polyubiquitin protein, heat shock protein 70 and an elongation factor 1 alpha.

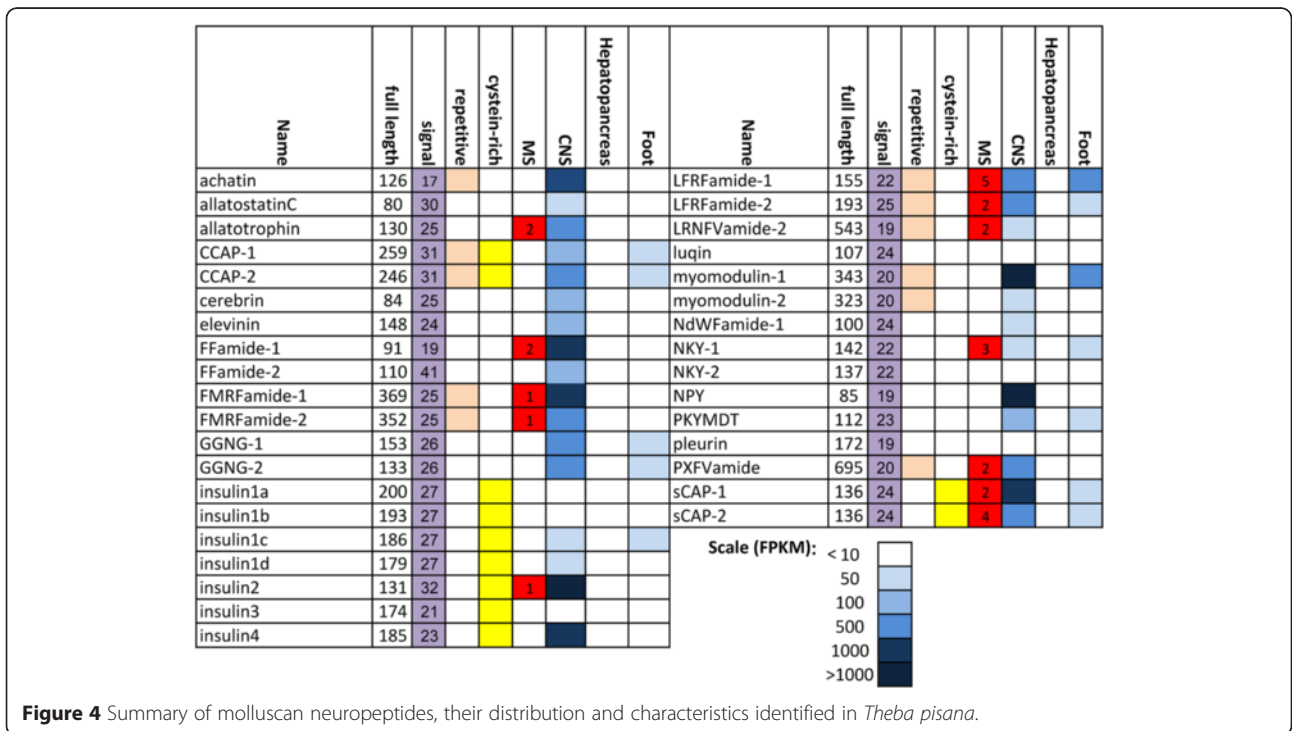
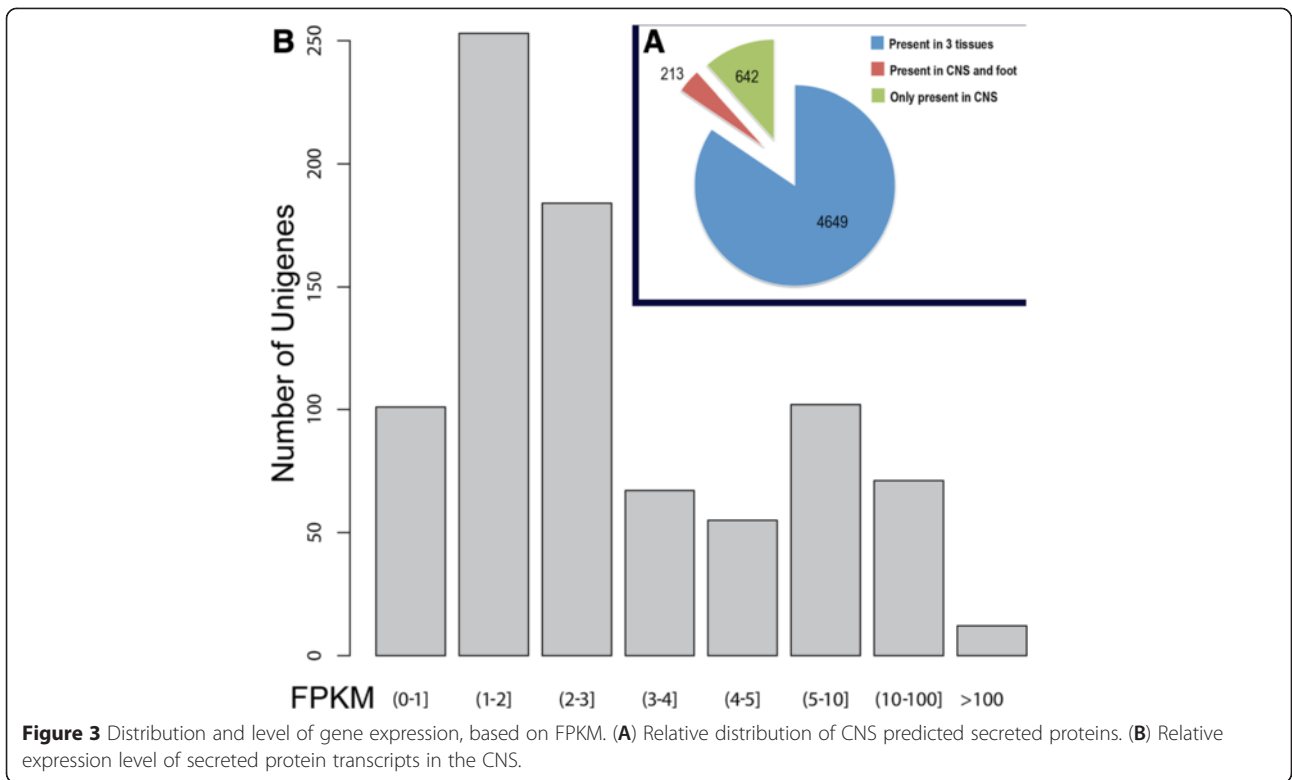
### Annotation of mollusc proteins secreted from CNS in *Theba pisana*

Investigation of the *T. pisana* CNS transcriptome, including identification of those precursors containing signal peptides and no transmembrane domains, revealed that *Theba* might contain as many as 5504 precursor proteins that are destined to secrete peptides (Additional file 3: Table S2). Of those, 4649 putative proteins are expressed in all three transcriptomes (Figure 3A). Another 213 were expressed in both CNS and foot, while there appeared to be no overlap in expression between CNS and hepatopancreas. The remaining 642 proteins were expressed exclusively in CNS. On the basis of a fold change higher than 5, we defined a list of 849 putative secreted proteins in CNS. As shown in Figure 3B, the expression FPKM for the majority of the putative secreted peptides were less than 3. Of those where expression was >100 and



10–100, 81% and 22% are designated as encoding hypothetical proteins, respectively (Additional file 3: Table S2).

Of the putative secreted proteins in the CNS, full-length precursors were identified in *T. pisana* for 35 known molluscan neuropeptides (Figure 4 and Additional file 4),





including isoforms and insulin that in molluscs can be secreted from neural tissue. In Figure 4, a “white” cell is defined as indicating only a trace amount or no transcript (<10). In general, for these neuropeptide transcripts, abundance was most prominent within neural tissue, particularly for insulin2, NPY, myomodulin1 and achatin (Figure 4 and Additional file 1: Figure S3).

Full-length precursors for some of these neuropeptides were also found to be of intermediate abundance in foot muscle while absent from the hepatopancreas. Peptide extraction followed by LC-MS analysis was used to confirm the usefulness of transcriptome-derived protein libraries and complement *in silico* neuropeptide predictions (Figure 4 and Additional file 1: Figure S4). Figure 5 shows the organisation of *T. pisana* neuropeptide precursors compared with previously identified homolog precursors of a marine gastropod, the limpet *Lottia gigantea*, with general high conservation in the spatial organisation of bioactive peptide sequences, as well as precursor size, cleavage sites and position of cysteine residues. Several of these neuropeptides were targeted for more in-depth tissue expression analysis.

The Tpi-Allatostatin C precursor comprised 80 amino acids, including signal peptide and cleavage sites to release the 15 amino acid active allatostatin C located in the C-terminal region of the precursor (Figure 6). Most similarity with Allatostatin C from other species was found within the bioactive peptide, primarily at the N-terminal SH and spatial conservation of two cysteine residues that likely form disulfide bonds to help stabilise the peptide. *Tpi-allatostatin C* was only identified in the CNS transcriptome, at a low expression level (Figure 4) but RT-PCR analysis showed that it was expressed not only in cerebral and pleural ganglia but also in *T. pisana* digestive tract and skin tissue (Figure 6). A representative RT-PCR of the 18 s rRNA gene is also shown, confirming the integrity of all tissue RNA. Two Crustacean cardioactive peptide (CCAP) precursor transcripts were identified in *T. pisana*, both encoding 3 CCAP moieties and showing very high amino acid identity with *Lottia* and *Aplysia* homologs (Figure 6). *Tpi-CCAP1* and 2 transcripts were identified in CNS and foot muscle transcriptomes, with CNS CCAP-2 being more abundant (Figure 4). RT-PCR demonstrated that *Tpi-CCAP1* is additionally expressed within sensory tentacles. The *Tpi-FFamide* transcript encoded 2 different amidated peptides; an LLFamide peptide and an LFFamide peptide, which show conservation with known homologs from other species within the C-terminal regions (Figure 6). Mass spectral analysis of the CNS of active *T. pisana* identified peptides corresponding to both LLFamide and LFFamide. Both *Tpi-FFa* transcripts were identified in the CNS transcriptome, with FF-1 very highly expressed (Figure 4). *Tpi-FFa-1* was additionally expressed within

the tentacles, digestive tract, skin and foot muscle tissues, despite not being identified from the foot muscle transcriptome.

Two *Tpi-NKYamide* transcripts were identified that encode neuropeptide Y (NKY) precursors of 142 and 137 residues, each containing cleavage sites to release two bioactive peptides, a 24-residue peptide and the 38-residue NKY peptide (Figure 6). Little amino acid conservation exists for the 24-residue peptide with other species. For the 38-residue NKY precursor, conservation with other species exists primarily within the N- and C-terminal regions, with very limited amino acid identity in the middle region. Mass spectral analysis of the CNS identified peptides corresponding to both NKY precursor peptides. Both *Tpi-NKY1* transcripts were observed in the CNS and foot muscle transcriptomes, while *Tpi-NKY2* was detected in very low amounts in the CNS transcriptome only. RT-PCR found that *Tpi-NKY1* was also present in the tentacles and *Tpi-NKY2* was additionally found in the tentacles, digestive tract, skin and foot muscle tissue. Two *Tpi-sCAP* transcripts were identified that both encode precursor proteins of 136 amino acids, including spatial conservation of putative cleavage sites and cysteine residues (Figure 6). High conservation between species was observed in the N-terminal region of the bioactive peptide. Mass spectral analysis of the CNS identified peptides corresponding to both sCAP precursor peptides. Both *Tpi-sCAP* transcripts were identified in high abundance within the CNS, and less in the foot muscle, while RT-PCR showed expression of *Tpi-sCAP1* in the tentacles, digestive tract and skin.

We identified from the three *T. pisana* transcriptomes, 7 transcripts with strong homology to insulins. Multiple sequence comparison and phylogeny demonstrated 4 groups of insulin-like peptides with strong bootstrap support; *Tpi-insulin1<sub>a-b</sub>*, *Tpi-insulin2*, *Tpi-insulin3* and *Tpi-insulin4* (Figure 4, Figure 7 and Additional file 1: Figure S5). *Tpi-insulin* precursors vary in size from the largest at 200 amino acids (*Tpi-insulin1<sub>a</sub>*) to the smallest at 131 amino acids (*Tpi-insulin2*). Consistent with other species, the organisation of *T. pisana* insulins consists of a “B peptide” with 3 cysteines, followed by a “C peptide” and then a C-terminal “A peptide” containing 5 cysteines. Only the *Tpi-insulin2* A peptide contains a glycine residue site for amidation. *Tpi-insulin2* and *Tpi-insulin4* were highly expressed in the CNS, while only *Tpi-insulin1<sub>c</sub>* could be identified in another transcriptome (foot muscle). No further RT-PCR analysis was performed to investigate insulin expression in other snail tissues.

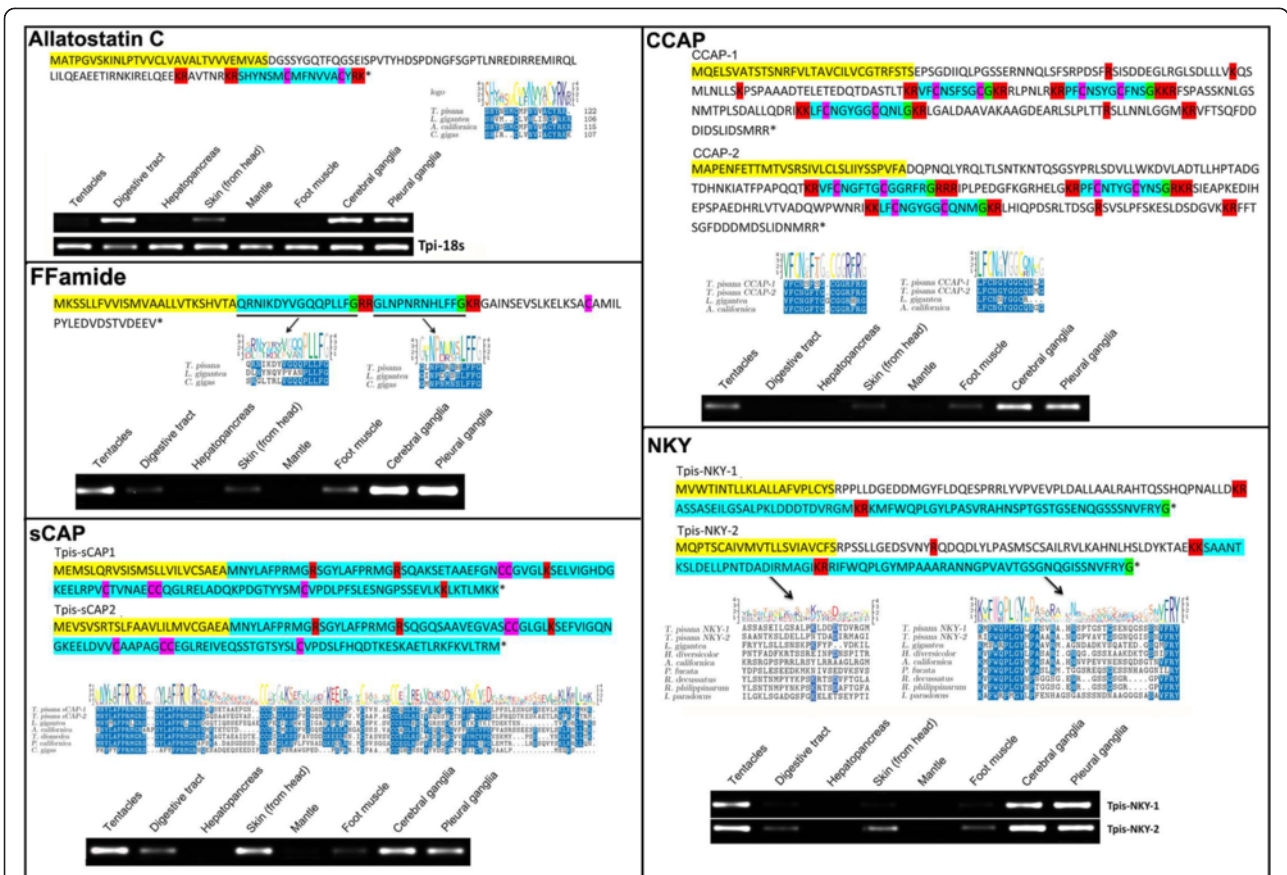
#### Prohormone convertases

We identified transcripts in each transcriptome database encoding enzymes that cleave precursor proteins, including the prohormone convertase 1 (PC1) and PC2 (Figure 8A, B

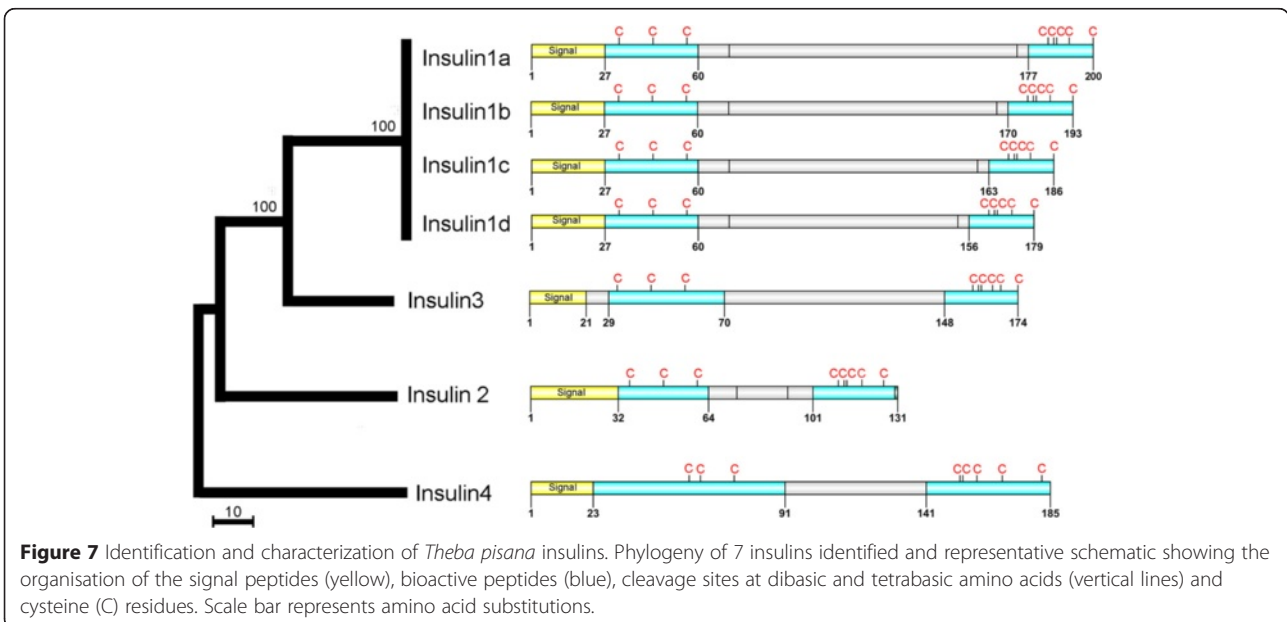


and Additional file 1: Figure S6). *Tpi-PC1* and *Tpi-PC2* encode precursors of 648 and 652 amino acids, respectively, and contain regions typical of these processing enzymes, including signal, pro and catalytic regions that are required for cleavage of dibasic amino acids.

Another type of convertase, the *Tpi-furin* was also identified that encodes a partial-length furin (554 amino acids) with a large 3' untranslated region. Multiple sequence alignment and phylogeny support their enzyme grouping (Figure 8C).

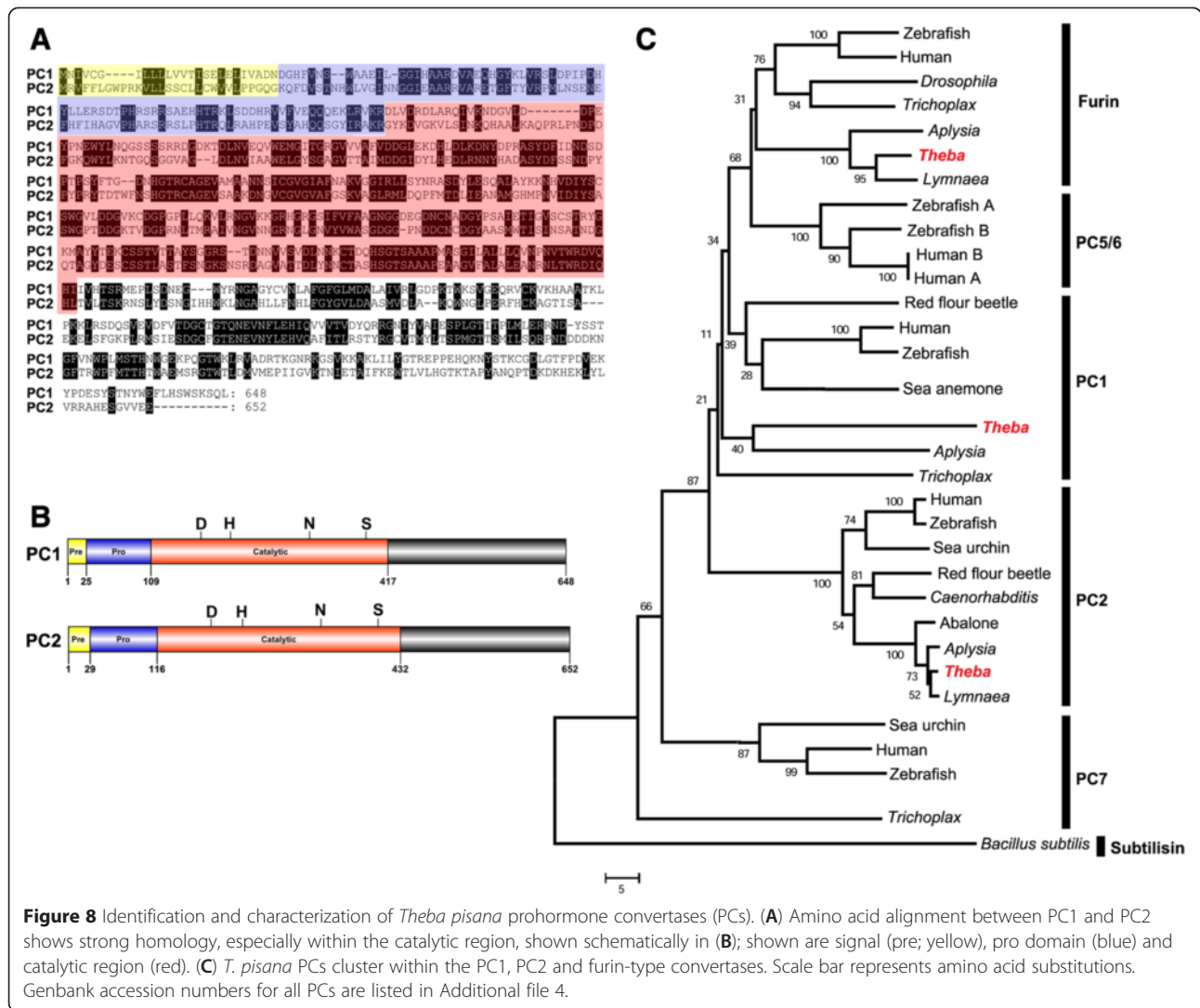


**Figure 6** Characterization of *Theba pisana* allatostatin C, CCAP, FFamide, sCAP and NKY. Each protein is represented by; (1) primary amino acid sequence with signal sequence (yellow), cleavage sites (red), cysteine residues (pink) and bioactive region (blue); (2) multiple amino acid sequence alignment with other gastropod and bivalve molluscs; and (3) RT-PCR expression of the neuropeptide gene within various tissues.



**Figure 7** Identification and characterization of *Theba pisana* insulins. Phylogeny of 7 insulins identified and representative schematic showing the organisation of the signal peptides (yellow), bioactive peptides (blue), cleavage sites at dibasic and tetrabasic amino acids (vertical lines) and cysteine (C) residues. Scale bar represents amino acid substitutions.





**Figure 8** Identification and characterization of *Theba pisana* prohormone convertases (PCs). **(A)** Amino acid alignment between PC1 and PC2 shows strong homology, especially within the catalytic region, shown schematically in **(B)**; shown are signal (pre; yellow), pro domain (blue) and catalytic region (red). **(C)** *T. pisana* PCs cluster within the PC1, PC2 and furin-type convertases. Scale bar represents amino acid substitutions. Genbank accession numbers for all PCs are listed in Additional file 4.

## Discussion

In this study, we extend on our understanding of gastropod biology by investigating transcriptomes of the land snail, *Theba pisana*. Deep transcript sequencing followed by *in silico* predictions and mass spectral analysis enabled the identification of numerous molluscan genes in this snail, including previously unknown neuropeptides. This greatly enhances our ability to utilise this species as an experimental model in neuroscience and metabolism, with the potential to enable development of molecular tools that could manipulate the physiology and behaviour of this agricultural and horticultural pest species.

Deep sequencing was performed to enable the assembly of three transcriptomes representing the CNS, hepatopancreas and foot muscle. Since gene expression levels may vary considerably depending on animal behavioural and physiological state, tissue samples for RNA transcriptome sequencing were combined from

stages representing active and aestivated, reproductive and juvenile snails. With over 300 million raw reads providing approximately 200,000 unigenes for each tissue, this ensured a comprehensive analysis of tissue genes. *De novo* assembly was necessary since no annotated reference is currently available for *Theba* or any other helicid snail. BLAST analysis showed that the majority of genes matched most closely with oyster, *Crassostrea gigas*, which represents the most expansive list of molluscan genes within the NCBI database due to the release of its genome and associated transcriptomes [18]. As a bivalve, *C. gigas* belongs to a different taxonomic group, thus our submission of the *Theba* transcripts derived from this study ensures that helicid snail gene sequences are now far better represented in the NCBI databases. Abundant unannotated transcripts suggest that there may be numerous helicid-specific sequences, or this species is at least considerably different to those currently annotated.

Gastropods have long been used in experimental neuroscience projects, providing a rich source for neurological discoveries. The vast majority of these studies have used marine and freshwater gastropods, including *Aplysia* and *Lymnaea*. In *Lymnaea stagnalis*, two separate studies analysed neural transcriptomes, first using normalized EST screening to identify 7,712 distinct transcripts [19], which was then expanded via Illumina sequencing to provide 116,355 contigs [20]. A large-scale neural EST screen of *Aplysia* CNS has revealed 175,000 transcripts, including 19,814 unique (at the time) neuronal gene products [21]. In our study, Illumina sequencing enabled a more comprehensive identification of *T. pisana* CNS transcripts, providing 220,602 unigenes. Several of the *Theba* neuropeptides identified have been shown to be significantly up- or down-regulated in the brain or hemolymph of hibernating *Helix pomatia* when compared to active individuals [15]. For instance, levels of peptides derived from the FMRFa, myomodulin, and probably LRNFVa precursors are much more abundant when *H. pomatia* are in the hibernation state, suggesting they may be critical for initiating and/or maintaining metabolic arrest. The CNS, hepatopancreas and foot muscle are crucial tissues regulating how snails initiate and maintain deep hypometabolic states, and therefore our comprehensive transcriptomes and associated transcriptome-derived protein databases should provide a useful platform to use for comprehensive gene-associated hypometabolic studies in these animals [22-24].

It has been speculated that in invertebrates, insulins, which are found within neurons, may have a role in neurotransmission [25]. Similar to vertebrate homologs, the invertebrate insulin precursor contains A and B chains, linked by a connecting (C) peptide. This C peptide is cleaved from the precursor and the A and B chains connect by cysteine bridges to form the mature insulin peptide [26]. In many cases, following cleavage of the precursor, the mature peptides are subjected to post-translational modification such as glycosylation, C-terminal amidation, acetylation, phosphorylation or sulfation [3]. Insulin precursors are known to undergo a series of processing events that yield the functional insulin, consisting of A and B peptides bound by intermolecular disulfide bonds. We found that *T. pisana* have 7 insulins, with each sharing similar attributes common to all other insulins. The largest reported molluscan insulin has been the *Aplysia* insulin; its involvement in regulating glucose metabolism was determined through the demonstration that *Aplysia* insulin can reduce hemolymph glucose levels [27]. Of all the *T. pisana* insulins, the *Tpi-insulin2* transcript appears to be the most highly expressed insulin in the CNS. Future studies are planned to determine the molecular role for this insulin.

This study shows that a prohormone convertase is present within a land snail. The ability for a cell to

secrete peptides relies on specific molecular secretory machinery, including processing enzymes that contribute to the synthesis of a mature bioactive peptide. In molluscs, PCs have been reported for *Lymnaea*, *Aplysia* and *Haliotis* [28-30]. The *Lymnaea* PC2 gene is exclusively expressed in the neuroendocrine system and may be present as two alternatively-spliced isoforms, similar to the *Haliotis* genes which are differentially expressed in neural tissues [28]. We found only one PC2 gene (as well as one PC1) in *T. pisana*, indicating that one of each is sufficient in this species to perform precursor cleavages, including cleavage of allatostatin C, FFamide and sCAP.

For gastropods, the hepatopancreas and foot muscle are major sites for lipid metabolism and storage, and therefore are critical during periods of starvation and hypometabolism [23,31]. The hepatopancreas is also a major source of digestive enzymes and is involved in nutrient absorption, food storage and excretion [32]. For that reason, it is not surprising that metabolic enzyme genes dominate transcripts annotated within the hepatopancreas. The foot muscle, on the other hand, contains numerous epiphragmin gene transcripts that encode for the epiphragmin protein, a major constituent of the epiphragm mucus that seals off the aperture during hibernation and aestivation [33]. The dried epiphragm is necessary to enable moisture retention while still allowing for gaseous exchange [34]. In the snail *Cerutuella virgata*, a major source of epiphragmin appears to be the mucous glands [33], while our study also implicates the foot muscle as a rich source of this protein. The foot contains its own type of mucus-secreting cells that were likely incorporated into our foot tissue transcriptome. During prolonged hypometabolism, muscle disuse atrophy (possibly related to reactive oxygen species) may be a major issue and we expect that future use of this transcriptome may enable us to establish molecular pathways by which these snails circumvent damage during hibernation and aestivation. Enzymes that have been implicated as key proteins in metabolic depression in other gastropod species, such as pyruvate kinase, phosphofructokinase and glutamate dehydrogenase (GDH) [35,36] are present in *T. pisana* transcriptomes and also represent obvious targets for future metabolic studies in this species. Furthermore, GDH activity can be as much as three times higher in aestivating snails than in active animals; GDH is crucial to amino acid metabolism and in aestivating snails helps to funnel nitrogen into urea biosynthesis which leads to the elevated urea in body fluids that contributes to desiccation resistance [22,37].

The identification of genes involved in the maintenance of land snails, as reported here, has numerous implications. Prior to this study, there was a lack of gene data for land snail species, despite their importance in

ecological functioning. *Theba pisana* is one of several snails native to the Mediterranean region that are now established as pests in grain crops, pastures and vineyards in southern Australia, where they not only eat seedling crops and pastures, but also clog harvesting equipment and contaminate grain harvests [38–40]. The current study provides a foundation for further studies into the manipulation of pest snail behavioural/physiological states that could lead to novel treatments to control their populations. For example, as demonstrated in insect pests [41,42], neuropeptide mimetic analogs hold potential for blocking or overstimulating receptors. There are also a number of other invasive snail species that have become major pests worldwide, including aquatic snail pests in reservoirs in Israel [43], the New Zealand mud snail *Potamopyrgus antipodarum* in Europe, Australia, Japan and North America [44], the golden apple snail, *Pomacea canaliculata*, in rice fields throughout Asia [45], and the giant African snail *Achatina fulica* throughout the tropics and subtropics [46]. In addition, the common garden snail *Cornu aspersum* Müller now inhabits large areas of Western Europe in man-made habitats [47] and is an introduced pest of many areas of the world, including North America [48], Australia [49] and New Zealand [50].

## Conclusion

This study has identified the genes expressed in three tissues of *Theba pisana* and identified numerous CNS neuropeptide gene products, confirming their expression *in vivo* by RT-PCR and mass spectrometry. Investigations can now progress into defining the function of gene products.

## Methods

### Animals and tissue collection

*Theba pisana* were collected from agricultural sites surrounding Warooka, located on the Yorke Peninsula, South Australia in early spring (September). Snails were transported to the University of the Sunshine Coast (USC) and housed within purpose-built enclosures. Snails were provided with water and food (cucumber and carrot) *ad libitum* and maintained at room temperature.

Active snails were conditioned by feeding and supplied with water daily over 7 days. Snails to be used for aestivation experiments were placed into glass jars without food or water and kept in an incubator on a cycle of 12 h at 30°C with light, 12 h at 20°C dark, to emulate South Australian summer conditions. The positions of the snails were marked on the jars after 14 days. Any snails that had not moved following a further 21 days were deemed to be in aestivation. To obtain waking snails, aestivated snails were sprayed with water and harvested as soon as they became active (10–60 min).

CNS, hepatopancreas (digestive gland) and foot muscle tissue was dissected out of *T. pisana* and immediately frozen. In each case, tissue from active, waking and aestivating snails was combined to maximise transcript representation from each metabolic state.

### RNA isolation and transcriptome sequencing

RNA was extracted from tissue using TRIzol Reagent (Invitrogen), as per the manufacturer's protocol. Following extraction, RNA was assessed for quality by visualisation on a 1.2% denaturing formaldehyde agarose gel, quantified using a Nanodrop spectrophotometer (Thermo scientific). For complimentary DNA (cDNA) synthesis, RNA samples were subjected to oligo-dT selection for mRNA purification and fragmented into small fragments. Fragmented RNA samples were subsequently repaired before adapter ligation. Suitable fragments were selected and reverse-transcribed into double-stranded cDNAs. The cDNA libraries were constructed by PCR amplification using random hexamer primed cDNAs. Finally, the samples were used for library construction and sequenced using an Illumina HiSeq 2000 sequencing (BGI, Hong Kong).

### Gene ontology and identification

*De novo* assemblies for each tissue type were performed by Trinity software [51] using trimmed reads from Illumina sequencing. The assembler was run with the parameter sets as following: seqType, fq; minimum kmer coverage = 4; minimum contig length of 200 bp. Sequences without Ns and which could not be extended on either end were defined as unigenes. When there were several samples from the same species, TGICL [52] was used to assemble all the unigenes from different samples to form a single set of non-redundant unigenes. After clustering, the unigenes were divided into two classes, clusters with the prefix CL, and singletons with the prefix Unigene.

Transcripts were annotated with the databases of NR, NT, Swiss-Prot, KEGG, COG, and GO, using BLAST and BLAST2GO software. Relative abundances of all transcripts among different tissues were estimated by SOAP software version 2.21. Searches of molluscan neurohormone precursors were also conducted by performing tBLASTn search of all tissue transcripts, which were reported by previous "omics" analysis studies [53–57]. BLAST searches were performed in CLC Main Workbench (Version 6.0). All hits were then analyzed manually with their orthologous peptides from various species and then their structures were characterized.

Analysis of protein identity/similarity between different protein receptors was performed by protein alignment using clustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). The percent identity was calculated as the number of identical amino acid residues, as indicated by "\*" symbol

in the Clustal output, divided by the total number of amino acid residues of the longest sequence ( $\times 100$ ). All hits (lowest E-value) were run through the SignalP 3.0 website for signal peptide prediction with the Neural Networks algorithm (Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark; <http://www.cbs.dtu.dk/services/SignalP/> ref). Proteolytic cleavage sites as well as post-translational modifications were predicted based on homology to other known peptides and the NeuroPred tool ([neuroproteomics.scs.illinois.edu/neuropred.html](http://neuroproteomics.scs.illinois.edu/neuropred.html)).

#### Protein comparison and Pfam domain annotation

To generate the most complete possible set of *T. pisana* peptides, we predicted protein-coding regions using Orf-Predictor [58] by default parameters on each tissue-specific assembly. We only retained the predicted longest ORFs and translated those into amino acid sequences over 30 amino acids. These sequences represented the full transcriptome-derived proteome. These were grouped into identical amino acid sequences using BLASTCLUST (BLAST score-based single-linkage clustering, [[ftp://ftp.ncbi.nih.gov/blast/documents/blastclust.html](http://ftp.ncbi.nih.gov/blast/documents/blastclust.html)]). We required the minimal length coverage threshold to be 0.7, which means that the minimum alignment length should cover at least 70% of full length of the shortest member in a group of sequences. For each resulting sequence clustering group, the longest amino acid sequence was chosen as the representative. The first optimal protein sequence in the sequence group was selected if multiple peptides had identical amino acid lengths. To provide an overview for the biological function, we annotated all the predicted proteins using the Pfam database (version 27.0). HMMSEARCH [42] was adopted to associate proteins with Pfam domains. We used a threshold of 0.01 for the e-value to identify reliable hits.

#### Neuropeptide prediction and sequence analysis

Neuropeptides are generally secreted out of the cell, which is facilitated by signal peptides in the premature protein form. To systematically identify putative neuropeptides in the CNS of *T. pisana*, we initially utilized four bioinformatics tools on all putative CNS proteins to predict the presence of a signal peptide (SignalP 3.0 [59] and PrediSi [60]) and any transmembrane domain (TMHMM 2.0 [61] and HMMTOP 2.1 [62]). For all these tools, we used default settings and parsed the results using in-house perl script. Then, resulting proteins were used as input to the NeuroPred tool to predict cleavage products. Schematic diagrams of protein domain structures were prepared using Domain Graph (DOG, version 2.0) software [63]. Protein sequences from *T. pisana* were aligned against a database prepared

from known sequences from NCBI (January, 2014) using the MEGA 5.1 [64] platform with the clustalW protocol utilising the Gonnet protein weight matrix. Neighbour-joining trees were generated based off these alignments. Unrooted trees were generated with 1000 bootstrap trials and presented with a cut-off bootstrapping value of 50.

#### Tissue distribution of selected genes of interest

Tissue was dissected out of 4 animals and each tissue type was pooled. Tissue types collected were tentacles, digestive tract, hepatopancreas, skin from head, mantle, foot muscle, cerebral ganglia and pleural ganglia. Total RNA was extracted from the 8 tissue types using TRIzol Reagent (Invitrogen, Catalogue # 15596-018) as per the manufacturer's protocol. RNA extracts were assessed using agarose gel electrophoresis, and quantified using a Nanodrop 2000 spectrophotometer (Thermo scientific). First-strand cDNA was synthesised using QuantiTect Reverse Transcription kit (Qiagen) as per the manufacturer's protocol. To normalise cDNA for each sample, equal quantities of extracted RNA were used as templates. Amplification of cDNA was carried out using the *Taq* F1 DNA Polymerase kit (Fisher Biotec) in a total volume of 25  $\mu$ l. The PCR reagent mix was prepared as recommended by the manufacturer, containing 1 $\times$  PCR reaction buffer, 2 mM  $MgCl_2$ , 0.2 mM each dNTPs, 0.2  $\mu$ M forward primer and reverse primer (sequences available upon request), 1 unit of *Taq* F1 DNA polymerase, 17.9  $\mu$ l water and 1  $\mu$ l cDNA template. A negative control, substituting water for template, was included in all PCR experiments. PCR was performed in a Kyratex Model SC200 thermal cycler. Cycling parameters were: initial denaturation at 95°C for 5 min, followed by 35 cycles of denaturation at 94°C for 20 s, annealing at 55°C for 30 s, and extension at 72°C for 1 min. Final extension was at 72°C for 10 min, then held at 10°C. Primers used were designed from sequences for *Theba pisana* 18 s rRNA gene and 6 peptide precursor genes (available upon request). A positive control reaction using cDNA from mixed tissues was also performed for each PCR reaction.

#### Peptide isolation from CNS and nanoHPLC, mass spectrometry LC-ESI-QTOF peptide identification

CNS were dissected out and combined from active and aestivated (3 weeks without movement) mature *T. pisana*, then immediately frozen in liquid nitrogen prior to storage at  $-80^\circ\text{C}$  until use. Frozen samples of CNS were ground to a powder under liquid nitrogen in a mortar, then quickly weighed and homogenized in extraction buffer (90% methanol, 9% glacial acetic acid in deionized water) in a 1:5 w:v ratio. Crude extracts were then sonicated with three pulses, 30 s each, and



centrifuged for 20 min (16,000 × g, 4°C). Supernatant was collected and lyophilised.

The CNS extracts were analyzed by LC-MS/MS on a Shimadzu Prominence Nano HPLC (Japan) coupled to a Triple-ToF 5600 mass spectrometer (ABSCIEX, Canada) equipped with a nano electrospray ion source. Aliquots (6 µl) of each extract were injected onto a 50 mm × 300 µm C18 trap column (Agilent Technologies, Australia) at 30 µl/min. The samples were desalted on the trap column for 5 minutes using solvent A [0.1% formic acid (aq)] at 30 µL/min. The trap column was then placed in-line with the analytical nano HPLC column, a 150 mm × 75 µm 300SBC18, 3.5 µm (Agilent Technologies) for mass spectrometry analysis. Peptide elution used a linear gradient of 1-40% solvent B [90:10 acetonitrile:0.1% formic acid (aq)] over 35 min at 300 nl/minute flow rate, followed by a steeper gradient from 40% to 80% solvent B over 5 min. Solvent B was then held at 80% for 5 min to wash the column and then returned to 1% solvent B for equilibration prior to the next sample injection. The ionspray voltage was set to 2400 V, declustering potential (DP) 100 V, curtain gas flow 25, nebuliser gas 1 (GS1) 12 and interface heater at 150°C. The mass spectrometer acquired 500 ms full scan TOF-MS data followed by 20 by 50 ms full scan product ion data in an Information Dependent Acquisition (IDA) mode. Full scan TOFMS data was acquired over the mass range 350–1800 and for product ion ms/ms 100–1800. Ions observed in the TOF-MS scan exceeding a threshold of 100 counts and a charge state of +2 to +5 were set to trigger the acquisition of product ion, ms/ms spectra of the resultant 20 most intense ions. The data was acquired and processed using Analyst TF 1.5.1 software (ABSCIEX, Canada).

Proteins were identified by database searching using PEAKS v7.0 (BSI, Canada) against the protein database built from the CNS transcriptome. Search parameters were as follows: no enzyme was used; variable modifications included methionine oxidation, conversion of glutamine to pyroglutamic acid, deamidation of asparagine and amidation. Precursor mass error tolerance was set to 20 ppm and a fragment ion mass error tolerance was set to 0.05 Da. Maximum expectation value for accepting individual peptide ion scores  $[-10 \cdot \log(p)]$  was set to  $\leq 0.05$ , where  $p$  is the probability that the observed match is a random event. Proteins and their supporting peptides were obtained and analysed.

#### Availability of supporting data

The raw reads of *T. pisana* mRNA are deposited in the NCBI GenBank as Sequence Read Archive (SRA) under the following accession numbers: SRP056280.

## Additional files

**Additional file 1: Figure S1.** Length distribution of unigenes. **Figure S2.** Graphs showing relative transcript abundance of top 50 transcripts in each *Theba pisana* transcriptome. **Figure S3.** Graph showing relative transcript abundance for *Theba pisana* neuropeptides identified. **Figure S4.** LC-MS/MS spectra showing peptides matching FFamide, NKY, sCAP and insulin precursors. **Figure S5.** *Theba pisana* insulin-like precursors annotated with features characteristic of insulin peptides. Yellow, signal peptide; blue, bioactive insulin; pink, cysteine residues; green, putative amidated glycine; red, cleavage sites. **Figure S6.** *Theba pisana* prohormone convertases, PC1 and PC2. Underline, signal peptide; red, catalytic region; boxed, conserved catalytic residues.

**Additional file 2: Table S1.** Transcriptome annotations for *Theba pisana* CNS, muscle and hepatopancreas.

**Additional file 3: Table S2.** Identification and annotation of secreted peptides within the CNS. All putative secreted peptides (Tab 1), gene ontology of secreted peptides (Tab 2), annotation of FPKM 10–100 (Tab 3) and FPKM >100 (Tab 4).

**Additional file 4: List of *Theba pisana* neuropeptides and prohormone convertases.**

#### Abbreviations

CNS: Central nervous system; COG: Cluster of orthologous groups; CCAP: Crustacean cardioactive peptide; FPKM: Fragments per kilobase of exon region in a given gene per million mapped fragments; GDH: Glutamate dehydrogenase; GO: Gene ontology; MS: Mass spectrometry; NKY: Neuropeptide Y; PC: Prohormone convertase; SCP: Small cardioactive peptide; USC: University of the Sunshine Coast.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

KJA carried out the bioinformatics and experimental analysis, constructed figures, tables and drafted the manuscript. MZ helped to perform the comparative transcriptome analysis. FB maintained animals and prepared tissue for experimental analysis. TW carried out mass spectral proteome work. AVK, KBS and SFC conceived the idea and obtained funding for the experiments and drafted the manuscript. All authors read and approved the final manuscript.

#### Acknowledgements

This work was supported by grants from the Australian Research Council (to SFC) and the Grains Research Development Corporation (to KA). We thank Dr. Alun Jones (Institute for Molecular Bioscience, the University of Queensland) for advice and assistance with the tandem mass spectrometry.

#### Author details

<sup>1</sup>Genecology Research Centre, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Maroochydore DC, Queensland 4558, Australia. <sup>2</sup>Institute of Biochemistry & Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada.

Received: 24 November 2014 Accepted: 31 March 2015

Published online: 17 April 2015

#### References

1. Benkendorff K. Molluscan biological and chemical diversity: secondary metabolites and medicinal resources produced by marine molluscs. *Biol Rev.* 2010;85(4):757–75.
2. Kiss T. Diversity and abundance: the basic properties of neuropeptide action in molluscs. *Gen Comp Endocrinol.* 2011;172(1):10–4.
3. Hökfelt T, Broberger C, Xu ZQD, Sergeev V, Ubink R, Diez M. Neuropeptides—an overview. *Neuropharmacology.* 2000;39(8):1337–56.
4. du Vigneaud V, Ressler C, Trippett S. The sequence of amino acids in oxytocin, with a proposal for the structure of oxytocin. *J Biol Chem.* 1953;205(2):949–57.



5. Popenoe EA, du Vigneaud V. A partial sequence of amino acids in performic acid-oxidized vasopressin. *J Biol Chem.* 1954;206(1):353–60.
6. Newcomb R, Scheller RH. Proteolytic processing of the Aplysia egg-laying hormone and R3-14 neuropeptide precursors. *J Neurosci.* 1987;7(3):854–63.
7. Painter SD. Coordination of reproductive activity in Aplysia: peptide neurohormones, neurotransmitters, and pheromones encoded by the egg-laying hormone family of genes. *Biol Bull.* 1992;183(1):165–72.
8. Chiu AY, Strumwasser F. An immunohistochemical study of the neuropeptidergic bag cells of Aplysia. *J Neurosci.* 1981;1(8):812–26.
9. Veenstra JA. Neurohormones and neuropeptides encoded by the genome of *Lottia gigantea*, with reference to other mollusks and insects. *Gen Comp Endocrinol.* 2010;167(1):86–103.
10. Weiss S, Goldberg JI, Chohan KS, Stell WK, Drummond GI, Lukowiak K. Evidence for FMRF-amide as a neurotransmitter in the gill of Aplysia californica. *J Neurosci.* 1984;4(8):1994–2000.
11. Harris LL, Lesser W, Ono JK. FMRFamide is endogenous to the Aplysia heart. *Cell Tissue Res.* 1995;282(2):331–41.
12. de Lange RPJ, van Golen FA, van Minnen J. Diversity in cell specific co-expression of four neuropeptide genes involved in control of male copulation behaviour in *Lymnaea stagnalis*. *Neuroscience.* 1997;78(1):289–99.
13. Smit A, Spijker S, Van Minnen J, Burke J, De Winter F, Van Elk R, et al. Expression and characterization of molluscan insulin-related peptide VII from the mollusc *Lymnaea stagnalis*. *Neuroscience.* 1996;70(2):589–96.
14. Stewart MJ, Favrel P, Rotgans BA, Wang T, Zhao M, Sohail M, et al. Neuropeptides encoded by the genomes of the Akoya pearl oyster *Pinctata fucata* and Pacific oyster *Crassostrea gigas*: a bioinformatic and peptidomic survey. *BMC Genomics.* 2014;15:840.
15. Pirger Z, Lubics A, Reglodi D, Laszlo Z, Mark L, Kiss T. Mass spectrometric analysis of activity-dependent changes of neuropeptide profile in the snail *Helix pomatia* Neuropeptides. 2010;44(6):475–83.
16. Artigaud S, Thorne MA, Richard J, Lavaud R, Jean F, Flye-Sainte-Marie J, et al. Deep sequencing of the mantle transcriptome of the great scallop *Pecten maximus*. *Mar Genomics.* 2014;15:3–4.
17. Pairett AN, Serb JM. De novo assembly and characterization of two transcriptomes reveal multiple light-mediated functions in the scallop eye (*Bivalvia: Pectinidae*). *PLoS One.* 2013;8(7): e69852.
18. Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature.* 2012;490(7418):49–54.
19. Feng Z, Zhang Z, Van Kesteren R, Straub V, Van Nierop P, Jin K, et al. Transcriptome analysis of the central nervous system of the mollusc *Lymnaea stagnalis*. *BMC Genomics.* 2009;10(1):451.
20. Sadamoto H, Takahashi H, Okada T, Kenmoku H, Toyota M, Asakawa Y. De novo sequencing and transcriptome analysis of the central nervous system of mollusc *Lymnaea stagnalis* by deep RNA sequencing. *PLoS One.* 2012;7(8): e42546.
21. Moroz LL, Edwards JR, Puthanveetil SV, Kohn AB, Ha T, Heyland A, et al. Neuronal transcriptome of Aplysia: neuronal compartments and circuitry. *Cell.* 2006;127(7):1453–67.
22. Bell RA, Dawson NJ, Storey KB. Insights into the in vivo regulation of glutamate dehydrogenase from the foot muscle of an estivating land snail. *Enzyme Res.* 2012;2012:17314.
23. Lama JL, Bell RA, Storey KB. Hexokinase regulation in the hepatopancreas and foot muscle of the anoxia-tolerant marine mollusc, *Littorina littorea*. *Comp Biochem Physiol B Biochem Mol Biol.* 2013;166(1):109–16.
24. Ito E, Okada R, Sakamoto Y, Otshuka E, Mita K, Okuta A, et al. Insulin and memory in *Lymnaea*. *Acta Biol Hung.* 2012;63 Suppl 2:194–201.
25. Boyd Jr FT, Clarke DW, Muther TF, Raizada MK. Insulin receptors and insulin modulation of norepinephrine uptake in neuronal cultures from rat brain. *J Biol Chem.* 1985;260(29):15880–4.
26. Huybrechts J, Bonhomme J, Minoli S, Prunier-Leterme N, Dombrovsky A, Abdel-Latif M, et al. Neuropeptide and neurohormone precursors in the pea aphid, *Acyrtosiphon pisum*. *Insect Mol Biol.* 2010;19:87–95.
27. Floyd PD, Li L, Rubakhin SS, Sweedler JV, Horn CC, Kupfermann I, et al. Insulin prohormone processing, distribution, and relation to metabolism in *Aplysia californica*. *J Neurosci.* 1999;19(18):7732–41.
28. Cummins SF, York PS, Hanna PJ, Degnan BM, Croll RP. Expression of prohormone convertase 2 and the generation of neuropeptides in the developing nervous system of the gastropod *Helicostoma*. *Int J Dev Biol.* 2012;53(7):1081–8.
29. Nagle GT, Garcia AT, Knock SL, Gorham EL, Van Heumen WR, Kurosky A. Molecular cloning, cDNA sequence, and localization of a prohormone convertase (PC2) from the Aplysia atrial gland. *DNA Cell Biol.* 1995;14(2):145–54.
30. Smit AB, Spijker S, Geraerts WP. Molluscan putative prohormone convertases: structural diversity in the central nervous system of *Lymnaea stagnalis*. *FEBS Lett.* 1992;312(2–3):213–8.
31. Boer M, Graeve M, Kattner G. Impact of feeding and starvation on the lipid metabolism of the Arctic pteropod *Clione limacina*. *J Exp Marine Biol Ecol.* 2006;328:98–112.
32. Barker GM. Mollusks as crop pests. Wallingford, Oxon, UK: CABI Publishing; 2002.
33. Li D, Graham LD. Epiphragmin, the major protein of epiphragm mucus from the vineyard snail, *Cermeuella virgata*. *Comp Biochem Physiol B Biochem Mol Biol.* 2007;148(2):192–200.
34. Barnhart MC. Gas permeability of the epiphragm of a terrestrial snail. *Otala lactea* Physiological Zoology. 1983;56(3):436–44.
35. Brooks SPJ, Storey KB. Properties of pyruvate dehydrogenase from the land snail, *Otala lactea*: control of enzyme activity during Estivation. *Physiol Zool.* 1992;65(3):620–33.
36. Ramnanan CJ, Storey KB. Glucose-6-phosphate dehydrogenase regulation during hypometabolism. *Biochem Biophys Res Commun.* 2006;339(1):7–16.
37. Bell RA, Storey KB. Regulation of liver glutamate dehydrogenase by reversible phosphorylation in a hibernating mammal. *Comp Biochem Physiol B Biochem Mol Biol.* 2010;157(3):310–6.
38. Baker G. The population dynamics of the mediterranean snails *Cermeuella virgata*, *Cochlicella acuta* (Hygromiidae) and *Theba pisana* (Helicidae) in pasture-cereal rotations in South Australia: a 20-year study. *Anim Prod Sci.* 2008;48(12):1514–22.
39. Baker GH. Recognising and responding to the influences of agriculture and other land-use practices on soil fauna in Australia. *Appl Soil Ecol.* 1998;9(1–3):303–10.
40. Baker G. Production of eggs and young snails by adult *Theba-Pisana* (Muller) and *Cermeuella-Virgata* (Da Costa) (Mollusca, Helicidae) in laboratory cultures and field populations. *Aust J Zool.* 1991;39(6):673–9.
41. Nachman RJ, Pietrantonio PV. Interaction of mimetic analogs of insect kinin neuropeptides with arthropod receptors. *Adv Exp Med Biol.* 2010;692:27–48.
42. Nachman RJ, Pietrantonio PV, Coast GM. Toward the development of novel pest management agents based upon insect kinin neuropeptide analogues. *Ann N Y Acad Sci.* 2009;1163:251–61.
43. Ben-Ami F, Heller J. Biological control of aquatic pest snails by the Black Carp *Mylopharyngodon piceus*. *Biol Control.* 2001;22(2):131–8.
44. Levri EP, Dermott RM, Lunnen SJ, Kelly AA, Ladson T. The distribution of the invasive New Zealand mud snail (*Potamopyrgus antipodarum*) in Lake Ontario. *Aquat Ecosyst Health Manag.* 2008;11(4):412–21.
45. Yusa Y, Sugiura N, Wada T. Predatory potential of freshwater animals on an invasive agricultural pest, the apple snail *Pomacea canaliculata* (Gastropoda: Ampullariidae), in southern Japan. *Biol Invasions.* 2006;8(2):137–47.
46. Thiengo SC, Faraco FA, Salgado NC, Cowie RH, Fernandez MA. Rapid spread of an invasive snail in South America: the giant African snail, *Achatina fulica*, in Brazil. *Biol Invasions.* 2007;9(6):693–702.
47. Madec L, Desbuquois C, Coutellec-vre MO. Phenotypic plasticity in reproductive traits: importance in the life history of *Helix aspersa* (Mollusca: Helicidae) in a recently colonized habitat. *Biol J Linn Soc.* 2000;69(1):25–39.
48. Selander RK, Kaufman DW. Genetic structure of populations of the brown snail (*Helix aspersa*). I. Microgeographic variation. *Evolution.* 1975;29(3):385–401.
49. Rudman B. Introduced snails in Australia. In: Australian Museum web site. 1999.
50. Barker GM, Watts C. Management of the invasive alien snail *Cantareus aspersus* on conservation land, Department of Conservation. 2002.
51. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29(7):644–52.
52. Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, et al. TIGER Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics.* 2003;19(5):651–2.
53. Ma M, Bors EK, Dickinson ES, Kwiatkowski MA, Sousa GL, Henry RP, et al. Characterization of the *Carcinus maenas* neuropeptidome by mass spectrometry and functional genomics. *Gen Comp Endocrinol.* 2009;161:320–34.
54. Gard AL, Lenz PH, Shaw JR, Christie AE. Identification of putative peptide paracrines/hormones in the water flea *Daphnia pulex* (Crustacea; Branchiopoda; Cladocera) using transcriptomics and immunohistochemistry. *Gen Comp Endocrinol.* 2009;160:271–87.
55. Christie AE, McCool MD, Harmon SM, Baer KN, Lenz PH. Genomic analyses of the *Daphnia pulex* peptidome. *Gen Comp Endocrinol.* 2011;171:131–50.
56. Veenstra JA, Rombauts S, Grbic M. In silico cloning of genes encoding neuropeptides, neurohormones and their putative G-protein coupled receptors in a spider mite. *Insect Biochem Mol Biol.* 2012;42:277–95.

57. Christie AE, Durkin CS, Hartline N, Ohno P, Lenz PH. Bioinformatic analyses of the publicly accessible crustacean expressed sequence tags (ESTs) reveal numerous novel neuropeptide-encoding precursor proteins, including ones from members of several little studied taxa. *Gen Comp Endocrinol.* 2010;167:164–78.
58. Min XJ, Butler G, Storms R, Tsang A. OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res.* 2005;33(Web Server issue):W677–80.
59. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol.* 2004;340(4):783–95.
60. Hiller K, Grote A, Scheer M, Munch R, Jahn D. PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Res.* 2004;32(Web Server issue):W375–9.
61. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001;305(3):567–80.
62. Tusnady GE, Simon I. The HMMTOP transmembrane topology prediction server. *Bioinformatics.* 2001;17(9):849–50.
63. Ren J, Wen L, Gao X, Jin C, Xue Y, Yao X. DOG 1.0: illustrator of protein domain structures. *Cell Res.* 2009;19:271–3.
64. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011;28(10):2731–9.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

