Journal of Proteomics 155 (2017) 22–30

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

An integrated proteomic and transcriptomic analysis of perivitelline fluid proteins in a freshwater gastropod laying aerial eggs

Huawei Mu [a], Jin Sun [b], Horacio Heras [c,d], Ka Hou Chu [e], Jian-Wen Qiu [a,*]

[a] Department of Biology, Hong Kong Baptist University, Hong Kong, China
[b] Division of Life Science, Hong Kong University of Science and Technology, Hong Kong, China
[c] Instituto de Investigaciones Bioquímicas de La Plata (INIBIOLP), Universidad Nacional de La Plata (UNLP)-CONICET CCT-La Plata, La Plata, Argentina
[d] Cátedra de Química Biológica, Facultad de Ciencias Naturales y Museo, UNLP, Argentina
[e] Simon F.S. Li Marine Science Laboratory, School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, China

ARTICLE INFO

Article history:
Received 22 October 2016
Received in revised form 4 January 2017
Accepted 8 January 2017
Available online 14 January 2017

Keywords:
Invertebrate
Egg
Apple Snail
Reproduction
Proteomics
Evolution

ABSTRACT

Proteins of the egg perivitelline fluid (PVF) that surrounds the embryo are critical for embryonic development in many animals, but little is known about their identities. Using an integrated proteomic and transcriptomic approach, we identified 64 proteins from the PVF of *Pomacea maculata*, a freshwater snail adopting aerial oviposition. Proteins were classified into eight functional groups: major multifunctional perivitellin subunits, immune response, energy metabolism, protein degradation, oxidation-reduction, signaling and binding, transcription and translation, and others. Comparison of gene expression levels between tissues showed that 22 PVF genes were exclusively expressed in albumen gland, the female organ that secretes PVF. Base substitution analysis of PVF and housekeeping genes between *P. maculata* and its closely related species *Pomacea canaliculata* showed that the reproductive proteins had a higher mean evolutionary rate. Predicted 3D structures of selected PVF proteins showed that some nonsynonymous substitutions are located at or near the binding regions that may affect protein function. The proteome and sequence divergence analysis revealed a substantial amount of maternal investment in embryonic nutrition and defense, and higher adaptive selective pressure on PVF protein-coding genes when compared with housekeeping genes, providing insight into the adaptations associated with the unusual reproductive strategy in these mollusks.

Significance: There has been great interest in studying reproduction-related proteins as such studies may not only answer fundamental questions about speciation and evolution, but also solve practical problems of animal infertility and pest outbreak. Our study has demonstrated the effectiveness of an integrated proteomic and transcriptomic approach in understanding the heavy maternal investment of proteins in the eggs of a non-model snail, and how the reproductive proteins may have evolved during the transition from laying underwater eggs to aerial eggs.

1. Introduction

There has been great interest over the last several decades in studying reproduction-related proteins as such studies not only have the potential to answer fundamental questions about speciation and evolution [1,2], but also to solve practical problems of animal infertility and pest outbreak [3,4]. In this regard, apple snails (Gastropoda: Ampullariidae) provide an excellent model for addressing numerous questions in evolutionary biology and good candidates to study these reproductive proteins due to a number of characteristics including high diversification of reproductive strategies [5]. In fact, while all members of Ampullariidae live in freshwater, their diverse oviposition strategies include laying egg masses either underwater, at the water surface or above the waterline [6]. This last strategy has seldom occurred in animals, and the egg proteins involved are the focus of the present study. In particular, species of the South American *Pomacea* deposit colorful egg masses above the waterline, which prevents the eggs from attack by aquatic predators, and is considered a critical innovation that has contributed to the diversification of *Pomacea* to become the most species-rich and derived genus of the family [6]. Nevertheless, embryonic development in aerial condition is challenging as the eggs are exposed to stressful abiotic conditions such as desiccation, heat and UV radiation, as well as terrestrial predators [7–9]. Therefore drastic structural and compositional changes in their eggs must have taken place to protect the embryos from abiotic stress and predation [10].

The *Pomacea* eggs contain a perivitelline fluid (PVF), which surrounds the embryo and provides energy and nutrients for embryonic

* Corresponding author.
E-mail address: qiujw@hkbu.edu.hk (J.-W. Qiu).

http://dx.doi.org/10.1016/j.jprot.2017.01.006
1874-3919/

development [10,11]. Proteins, termed perivitellins, are an important component of organic matter in the PVF, ranging from 13% up to 18% of the total dry weight [11,12]. Previous reports in the best studied *Pomacea* species, *Pomacea canaliculata* (Lamarck, 1822), characterized the two most abundant perivitellins (termed PcOvo and PcPV2) and found that they not only serve as a source of energy and nutrient, but also play several defensive functions against abiotic stressors and predators [9,13–18]. Moreover, a proteomic study identified 59 proteins in the PVF of *P. canaliculata*, showing that the PVF proteome is complex and highlighting the need of more studies on *Pomacea* reproductive proteins to better understand their adaptations [10].

Comparison between closely related species is an effective approach to study the evolution of reproductive proteins [19–21]. In this regard, *Pomacea maculata* (Perry, 1810), a close relative of *P. canaliculata* [22], is a good candidate to address the PVF protein evolution. The two species have been widely introduced out of South America in tropical and subtropical North America [23] and Asia [24–26] becoming invasive pests of agricultural and natural wetlands. Their distribution ranges overlap in their native South American and invaded Asian wetlands [23,24] and they can hybridize naturally [27]. Nevertheless, the two species have different distribution ranges associated with their differential cold tolerance [27]. Therefore comparing the PVF proteins between the two species has the potential to reveal their contribution to adaptation and further range expansion. Nevertheless, little is known about the perivitellins of *P. maculata*. Recent studies have reported two major perivitellins of *P. maculata*, which are structurally and functionally similar to the corresponding PVF proteins of *P. canaliculata* [12,28]. However, very little is known about the presence and roles of the less abundant PVF proteins of *P. maculata* [12].

We therefore profiled the PVF proteome of *P. maculata* by using a combined *de novo* transcriptome assembly and LC-MS/MS approach. We hypothesized that perivitellin genes would be actively expressed in albumen gland (AG) when compared with genes coding non-PVF proteins. We therefore compared the gene expression level of PVF proteins in AG against that of other tissues. In addition, we analyzed the substitution rates of protein-coding genes between *P. maculata* and *P. canaliculata* in order to gain insight into the mechanisms of their divergence.

2. Materials & methods

2.1. Snail culture

Adults of *Pomacea maculata* (>30 mm shell length) were originally collected from the Paraná River near San Pedro, Argentina (33°39′ 35.97″ S, 59°41′52.86″ W), cultured at Universidad Nacional de La Plata (Argentina) and later transported to Hong Kong Baptist University. The identity of *P. maculata* was confirmed by amplifying fragments of the elongation factor 1-alpha (EF1α) and cytochrome *c* oxidase subunit I (COI) genes [27] and comparing them with the sequences deposited in GenBank. Snails were cultured at 25 ± 1 °C in 250-L tanks filled with tap water. Aeration was provided by an air stone connected to an air pump, and the water was cleaned using an overflow filter. Tap water used in the snail culture was dechlorinated by aeration before use, and changed once a week. The snails were fed with lettuce, carrot and fish feed. Food leftovers and snail feces were removed daily. Under these conditions, the snails were able to grow, mate and deposit egg clutches on the tank wall.

2.2. Transcriptome sequencing, assembly and annotation

To obtain a database for protein identification following MS/MS analysis, and to detect genes specifically expressed in the AG, we sequenced the transcriptomes of AG, and other tissues (OT: foot, mantle and visceral mass) pooled in equal amount from four individuals (Fig. 1). Total RNA from AG and OT was extracted separately using the TRIzol reagent (Invitrogen, Carlsbad, USA) following the manufacturer's manual except that a high salt solution (mixed solution of 1.2 M sodium chloride and 0.8 M sodium citrate) and a lithium chloride solution (final concentration 2 M) were added before and after the isopropanol precipitation step to remove polysaccharides. RNA quality was checked using agarose gel electrophoresis and an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The RIN value for the albumen gland and other tissues sample was 8.1 and 9.1, respectively. Messenger RNA was selected, reverse-transcribed into cDNA, and paired-end sequenced on an Illumina Hiseq 2000 to produce 100 base pair (bp) reads (Illumina, San Diego, CA, USA). Reads were assembled using Trinity (release 20130225) [29]. The generated sequences were annotated by searching against protein databases (NCBI nr, KEGG, COG, and Swissprot) using BLASTx with an *E*-value threshold of $1 \times e^{-5}$ [30]. Protein-coding regions were translated into amino acid sequences and used as a database for protein identification.

2.3. Egg mass collection and protein extraction

Egg masses laid within 12 h were carefully removed from the culture tank wall, rinsed several times with MilliQ water, and then air-dried in a fume hood. When the egg shells were hardened and became non-sticky, each egg was individually cracked gently using a fine sterile needle and the perivitelline fluid (PVF) was extracted using 10 μL pipette tips. The PVF was dissolved in 8 M urea, homogenized thoroughly and centrifuged at 12,000 ×*g* for 10 min at 4 °C. Then the supernatant was transferred into new tubes. The protein solution was purified using the methanol/chloroform method [31]. Three biological samples were collected from different egg masses coming from different females, and the protein concentrations were measured with the RC-DC kit (Bio-Rad).

2.4. SDS-PAGE and LC–MS/MS analysis

The purified protein solution was mixed with the SDS-PAGE buffer (50% glycerol, 0.2 M Tris-HCl pH 6.8, 0.05% bromophenol blue, 10 mM dithiothreitol, and 10% SDS) at a ratio of 3:1 (v/v), and placed on a heating block at 105 °C for 5 min. Proteins in each sample were separated by SDS-PAGE and stained by Coomassie Brilliant Blue. Then 1% acetic acid was applied to destain the gel. Each sample was divided into 10 slices based on their band intensity and protein molecular weight (Fig. 1).

Gel slices were further destained with a solution of 50% methanol/50 mM NH_4HCO_3, washed with MilliQ, dried with 100% ACN, and re-hydrated in 100 mM NH_4HCO_3. Dithiothreitol (10 mM) and iodoacetamide (55 mM) were sequentially applied to reduce the disulfide bonds of proteins and alkylate the exposed sulfhydryl (−SH) groups, respectively. The gel was then digested with sequencing grade trypsin (Promega, Madison, WI) in 50 mM NH_4HCO_3. Peptide solutions were recovered from the gel, desalted using Sep-Pak C18 cartridges (Waters, Milford, MA), and dried in a vacuum concentrator (Eppendorf, Hamburg, Germany).

The dried fraction of each biological sample was reconstituted with 0.1% formic acid and analyzed twice with a LTQ-Orbitrap Elite coupled to an Easy-nLC (Thermo Fisher, Bremen, Germany) as described [10]. Briefly, peptides were separated in a C18 capillary column (Michrom BioResources, CA) using a 90-min gradient. Mass spectrometry scans with a range of 350 to 1600 *m/z* were acquired with a resolution of 60,000 under the positive charge mode. The five most abundant multiple-charged ions having a minimum signal threshold of 500.0 were selected for high-energy collision-induced dissociation (HCD) and fragmentation using collision-induced dissociation (CID). Both the HCD and CID scanning strategies adopted an isolation width of 2.0 *m/z*. For HCD fragmentation, the activation time was 10 ms and the normalized

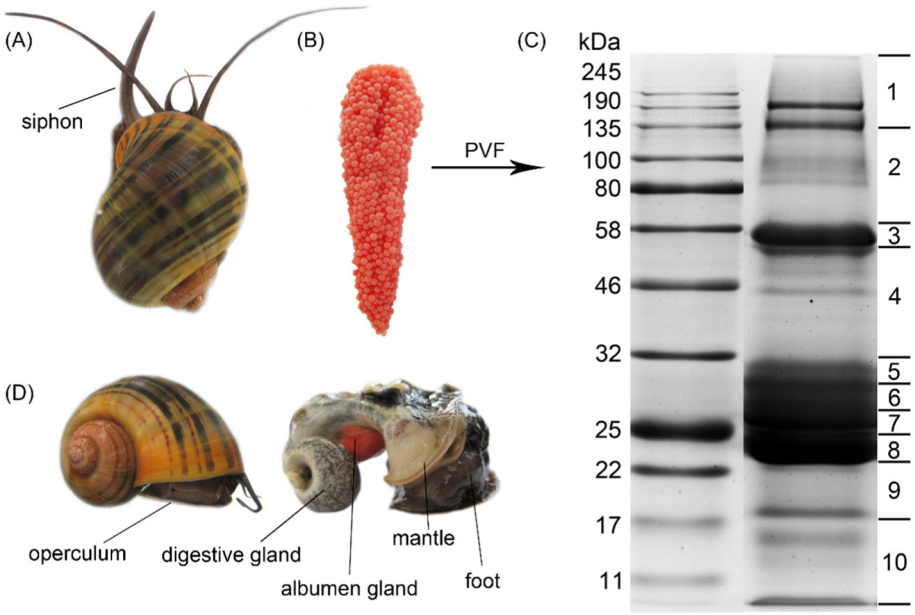

Fig. 1. *Pomacea maculata.* An individual (shell length = 63 mm) crawling near the water surface (A). A clutch of newly deposited eggs (length = 7.9 cm) (B). A representative SDS-PAGE of the egg perivitelline fluid proteins (right lane) and protein molecular weight markers (kDa, left lane) (C). Numbers 1 to 10 on the right indicate the gel slices used for LC-MS/MS analysis. Anatomy of *P. maculata* with and without shell showing the positions of several structures and organs (D).

collision energy was 45%. For CID fragmentation, the activation time was also 10 ms, but the normalized collision energy was 35%.

2.5. Database search

Raw LC-MS/MS data files were converted into *.mgf* files using Proteome Discovery 1.3.0.339 (Thermo Finnigan, CA), and submitted to Mascot version 2.3.2 (Matrix Sciences, London, U.K.) to search against the *P. maculata* database with 77,584 protein sequences including both 'target' and 'decoy' sequences. Searching parameters were identical to those described in Mu et al. (2015) [32] except that only one maximum missed cleavage of trypsin was allowed, and fixed modification was set for cysteine carbamidomethylation. Matched peptides with an ion score ≥ 22 were retained to achieve a 95% confidence in identification. Only peptides longer than nine amino acids were retained because shorter peptides could easily match the decoy database. A 1% false discovery rate threshold was applied in the final protein identification in each biological replicate. Only proteins detected in at least two replicates and had at least three matched peptides were kept.

2.6. Transcriptomic expression analysis

To distinguish genes that were specifically expressed in AG and OT, we compared their expression levels quantified as transcripts per million (TPM) which takes both gene length and sequence depth into account [33]. TPM was calculated by dividing the read counts by the length of each gene in kilobases to get a reads per kilobase (RPK) value, summing up all the RPK values in AG and OT and dividing this number by 1,000,000 to get a "per million" scaling factor, then dividing the RPK values by the scaling factor. Genes with an expression level (TPM) less than 0.5 were considered to be non-expressed [29].

Reverse transcription-PCR (RT-PCR) reactions were performed to determine the tissue specific expression pattern of seven PVF protein-coding genes. RNA samples from albumen gland, digestive gland, foot, mantle, and testis were reverse transcribed into complementary DNA (cDNA) using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems), then the cDNAs were used as templates in PCR experiments to show gene expression. Cytochrome *c* oxidase I (COI) was chosen as control as this gene had high expression levels in the different tissues, and has been considered as a house keeping gene [34]. Primers were designed using NCBI's Primer Designing Tool and their sequences are shown in Supplementary Table S1. PCR reactions were performed in 20 μL solution containing 0.2 μL of TaKaRa Taq™ DNA Polymerase, 1 μL of primers (final working concentration 0.5 μM), 1 μL of DNA template, 2 μL of 10 × reaction buffer, 1.6 μL of dNTP (final working concentration 0.2 mM), and 1.2 μL of 25 mM of $MgCl_2$ and 13 μL of DNase-free water. PCR conditions were one cycle of 5 min at 94 °C, 25 cycles of 30 s at 94 °C, 40 s at 57 °C and 40 s at 72 °C, followed by a 10 min extension at 72 °C, and a final 30 min incubation at 4 °C. Due to the technical difficulty of having no antibodies for the non-model organism, we were not able to perform Western Blot to confirm the protein expression levels.

2.7. Evolutionary analysis

A reciprocal best hit method was applied to identify ortholog pairs between *P. maculata* and *P. canaliculata* transcriptomes [32,35]. Only proteins with at least 67 amino acids were used in the orthologous analysis. Local BLASTp was performed with an *E*-value threshold of $1 \times e^{-5}$. Paired orthologs between the two species were aligned by MUSCLE implemented in ParaAT1.0 [36] and aligned codons with gaps were removed. KaKs_Calculator was applied to calculate the nonsynonymous substitutions per nonsynonymous site (Ka), synonymous substitutions per synonymous site (Ks), and Ka/Ks ratio using the GY method [37]. Sequences with Ka > 1 or Ks > 1 were removed because they might be saturated for nonsynonymous or synonymous substitutions, respectively [38]. Genes with a Ks < 0.01 were excluded in order to avoid pseudo-high Ka/Ks values due to the low Ks value [39]. To determine whether the genes coding PVF proteins evolved faster than conserved genes, the Ka, Ks, and Ka/Ks values of PVF gene orthologs were compared with three groups of housekeeping genes: actin, tubulin and ribosomal proteins (Supplementary Table S2) [40] using the non-parametric Wilcoxon rank sum test [40,41].

2.8. Bioinformatic analysis

Protein abundance was quantified using the Exponentially Modified Protein Abundance Index (emPAI) in Mascot [42]. The PVF proteins were classified into functional categories based on their annotation. SignalP 4.1 was applied to determine whether the PVF proteins were secreted [43]. For PVF proteins with a Ka/Ks value > 0.5, their three-dimensional structures were predicted using Phyre2 which applied a profile-profile alignment algorithm [44]. Models having a confidence level of >90% in Phyre2 were validated using PROCHECK and those having >90% residues in the most favored and additional allowed regions were kept. The potential ligand binding sites (or active sites) were then predicted by 3DLigandSite [45]. Sequences were also searched against the Conserved Domain Database to detect the conserved protein regions/domains [46].

3. Results

Transcriptome sequencing produced a total of 52,732,156 and 54,961,478 clean reads for AG and OT, respectively. The mean GC content of AG and OT sequences was 44.94% and 45.05%, respectively. Assembly of the sequences using Trinity produced 92,567 and 130,305 unigenes in AG and OT, respectively. Combining the unigenes from the two organs resulted in 105,349 unigenes and the N50 is 1332 nt. Among them, 38,587 were successfully annotated by searching against several protein databases (i.e., NCBI nr, SwissProt, KEGG, GO and COG). These annotated unigenes were used for protein identification.

With the MS data and the aid of the assembled transcriptome database, we identified 64 proteins in the PVF of *P. maculata* based on 6995 peptides, among which 364 peptides were unique (Table 1). The PVF proteins were classified into eight functional categories: major multifunctional perivitellin subunits (13), immune responsive proteins (14), oxidation-reduction (3), signaling and binding (10), protein degradation (7), energy metabolism (4), transcription and translation (4), and others (9). The emPAI method showed that the abundance of these 64 proteins varied over a broad range from 0.02% to 24.62%, indicating that our analytical approach was sensitive enough to detect even low-abundance proteins. The six most abundant PVF proteins were PmPV1–2 (24.62%), PmPV1-4a (18.90%), PmPV1-4b (14.28%), perivitellin ovorubin-3 (8.49%), perivitellin ovorubin-2 (7.32%) and PmPV1-1 (3.83%). Immune responsive proteins which played a vital role in egg defense were also abundant as the 14 proteins accounted for 7.46% of the total PVF proteins. The three less abundant functional groups of proteins were signaling and binding (2.20%), protein degradation (1.33%) and energy metabolism (0.96%). There were also seven ¨novel¨ proteins which were not functionally annotated.

Since AG is considered the organ that synthesizes most of the PVF proteins in *P. canaliculata* [10], the expression of genes coding the PVF proteins in AG is expected to be higher than in OT. Indeed, comparison of AG and OT transcriptomes showed that genes coding 22 of the PVF proteins were only expressed in AG, and as the majority of PVF proteins were preferentially expressed in AG (Table 1). The five most highly expressed PVF genes in AG were PmPV1-4b, perivitellin ovorubin-2, PmPV1-4a, PmPV1-2, and perivitellin ovorubin-3, all with a TPM > 43000. Their expression levels in OT were all less than 0.3, below the threshold of 0.5 considered to be non-expressed in a previous study [29]. Seven PVF protein-coding genes were exclusively expressed in AG (Fig. 2). Fig. 3 shows the correlation between the abundances of 64 PVF proteins and their corresponding gene expression levels in AG and OT. Protein abundance was not significantly correlated with gene expression level in OT; but was positively correlated with gene expression level in AG, consistent with the fact that AG is the organ where most of the PVF proteins are synthesized.

Base substitution analysis of the orthologs between the two species revealed 12,231 pairs of genes having Ka/Ks values. Among the 64 PVF gene coding proteins, 30 had Ka/Ks values (Supplementary Table S2). Five genes had a Ka/Ks value > 1, and six genes had a Ka/Ks between 0.5 and 1 (Fig. 4). Comparison between PVF genes and housekeeping genes showed that the PVF genes had significantly higher mean Ka (0.033 vs. 0.003), Ks (0.081 vs. 0.058) and Ka/Ks (0.476 vs. 0.072) values (Supplementary Fig. S1; Wilcoxon rank sum test, $P < 0.05$).

Homology modeling of the 11 proteins having Ka/Ks > 0.5 using Phyre2 with a threshold of 90% confidence showed that, in 5 of them, the predicted 3D structure had a reasonable match to the corresponding template (Fig. 5, Supplementary Fig. S2). Analysis with 3DLigandSite revealed that some of the nonsynonymous substitution sites and potential ligand binding/active sites overlapped or were very closely located (e.g., Kunitz-like protease inhibitor, ferritin (partial), and transferrin-like protein), indicating the potential of nonsynonymous substitution to modify protein conformation and thus their function. For instance, in ferritin (partial) there were 115 amino acid residues matching its homologous template (PDB ID: 3VNX) with a relatively high sequence similarity (49%). The confidence that this snail ferritin and its template are homologous was 100%, indicating it was reasonable to predict the 3D structure of the PVF protein based on that of 3VNX. As shown in Fig. 5, the snail ferritin has four alpha helixes (76% of amino acids) and one short beta strand (3% of amino acids). The molecule has three predicted ligand binding sites (8E, 11H and 53E), four putative ferrihydrite nucleation centers (3K, 6D, 7E and 10E) and three putative iron ion channels (64H, 77N and 80E). Comparison with its ortholog of *P. canaliculata* revealed 11 nonsynonymous substitution sites, some overlapping either with the binding site or the nucleation centers of ion channels. The homology models of the other proteins are depicted in Fig. S2.

4. Discussion

4.1. Functional composition of PVF proteins

The PVF of *P. maculata* eggs contain 64 proteins belonging to many functional groups, indicating a heavy maternal investment to support the embryonic development. The most abundant *P. maculata* PVF proteins are the subunits of the two major perivitellins, namely PmPV1 and PmPV2 [28]. A previous study on the sister species *P. canaliculata* indicates that the orthologs of these perivitellins are an energy source and structural precursors of other biological molecules during embryonic development [11]. PmPV1, representing ~52% of PVF protein, is a glyco-lipo-carotenoprotein responsible for the bright red coloration (presumably a warning signal) and embryo protection from oxidative damage [28].

Among the 13 *P. maculata* multifunctional perivitellin subunits detected, the most abundant group is the PmPV1 subunits, which is in agreement with a previous study of *P. canaliculata*, where subunits of PcOvo, the ortholog of PmPV1, are also the most abundant [10]. Interestingly, the 5 subunits differed markedly in their relative abundance, pointing to the potential of different arrangements in the native protein structure in the two species. For example, the PmPV1-2 has an emPAI of 24.41%, but its corresponding ortholog in *P. canaliculata* is 48.61%. The second most abundant perivitellin group in *P. maculata* is PV2 which is antinutritive, antidigestive and neurotoxic in *P. canaliculata*, due to the presence of a light chain with lectin binding property and a heavy chain with perforin properties [9]. Sequence alignment of the full amino acid sequences of the light and heavy chain of PV2 (i.e, tachylectin and MACPF subunits) [9,10] between *P. maculata* and *P. canaliculata* showed 97% and 96% similarities, respectively (Supplementary Fig. S3). The nonsynonymous substitution sites (9 and 20 amino acid sites in tachylectin and MACPF subunit, respectively) were not located in the phosphorylation, glycosylation sites or MACPF signature regions. This was consistent with a previous study on PcPV2 [9] showing that the MACPF signature in *Pomacea* was quite conserved, which might be related to their adaptation to the aerial egg deposition strategy as an egg defense protein. A comparison of PV3 between the two species showed that they differed markedly both in protein abundance and

Table 1
Proteins identified from the egg perivitelline fluid of *Pomacea maculata*.

Name	Annotation	MW (KDa)	No. of peptides	Average %[a]	Pm_TPM[b]	
					AG	OT
Major multifunctional perivitellin subunits						
CL7293.C3	PmPV1-2[c]	22.499	770	24.62 ± 3.65	**46,075.76**	0.15
Unigene35350	PmPV1-4a[c]	21.944	1148	18.90 ± 8.96	**66,379.16**	0.10
Unigene40233	PmPV1-4b[c]	22.368	800	14.28 ± 3.50	**92,663.06**	0.26
Unigene34885	Perivitellin ovorubin-3[c]	22.68	226	8.49 ± 1.49	**43,994.87**	0.10
Unigene34956	Perivitellin ovorubin-2[c]	23.748	385	7.32 ± 0.74	**80,798.28**	0.00
CL2290.C1	PmPV1–1[c]	23.226	238	3.83 ± 1.35	**3889.26**	0.00
CL6258.C4	PV2 MACPF subunit[c]	63.171	674	1.77 ± 0.37	4747.31	0.55
CL7293.C1	perivitellin protein	8.951	52	0.76 ± 0.12	**15,012.89**	0.00
CL4382.C1	PV2 tachylectin subunit[c]	31.728	180	1.17 ± 0.38	**1979.86**	0.45
CL4382.C2	PV2 tachylectin subunit[c]	32.16	191	0.72 ± 0.21	1465.04	1.25
CL5816.C3	Perivitellin protein[c]	22.491	95	0.67 ± 0.22	**11.51**	0.00
Unigene34835	Tachylectin-like protein	29.536	80	0.59 ± 0.27	**2580.39**	0.09
CL5816.C1	Perivitellin protein[c]	22.505	80	0.40 ± 0.16	**40.45**	0.00
Immune responsive proteins						
Unigene40434	Ferritin, partial	13.144	24	2.09 ± 0.46	**1090.93**	0.39
CL9552.C2	Transferrin-like protein	36.762	79	0.84 ± 0.09	53.68	1.19
Unigene40228	C1q domain-containing protein[c]	18.338	53	0.63 ± 0.22	**19,602.56**	0.06
CL9552.C3	Transferrin-like protein	81.962	95	0.55 ± 0.15	56.52	1.64
Unigene38276	Serine protease inhibitor-1, partial[c]	8.144	23	0.49 ± 0.24	122.04	0.52
CL6429.C2	Serpin 1	39.27	95	0.45 ± 0.05	**4856.93**	0.03
CL8110.C1	Peptidoglycan recognition protein S1L	11.343	24	0.42 ± 0.17	**51.85**	0.00
CL3024.C1	Universal stress protein Slr1101-like	14.563	3	0.41 ± 0.27	6.70	8.26
CL539.C1	Deleted in malignant brain tumors 1 protein	11.324	183	0.35 ± 0.18	796.74	0.76
CL1118.C4	Kunitz-like protease inhibitor[c]	21.582	107	0.33 ± 0.09	**1938.09**	0.30
CL3601.C1	Thioester-containing protein-C	157.143	127	0.27 ± 0.10	39.41	37.61
CL4196.C2	Alpha-2-macroglobulin	31.204	18	0.20 ± 0.05	2.89	3.01
Unigene28932	Ferritin, partial	19.276	3	0.21 ± 0.13	134.29	174.65
Unigene35143	CD109 antigen-like	43.83	21	0.22 ± 0.06	3.95	11.95
Oxidation-reduction						
CL9797.C1	Pi-class glutathione S-transferase	24.322	8	0.15 ± 0.10	41.03	12.27
CL4528.C1	Putative ferric-chelate reductase 1 homolog	62.479	30	0.10 ± 0.01	7.21	1.64
Unigene7542	Superoxide dismutase [Cu-Zn]	103.473	19	0.07 ± 0.01	88.14	22.51
Signaling and binding						
CL762.C1	Calcium-binding protein, partial [c]	17.683	169	0.63 ± 0.14	10,530.26	0.89
Unigene31436	PREDICTED: ras-related protein Rab-10-like	22.096	32	0.27 ± 0.14	14.19	30.96
CL762.C2	Calcium-binding protein, partial	14.856	5	0.46 ± 0.18	15,846.13	6.41
Unigene14225	Ras-like protein	21.046	14	0.23 ± 0.04	26.77	57.64
Unigene33406	RAB protein	23.054	19	0.16 ± 0.08	9.29	30.81
CL1720.C2	RAB3	20.738	8	0.21 ± 0.10	2.33	4.16
Unigene3919	Ras-related protein Rab-33B	26.752	16	0.06 ± 0.03	4.02	4.08
CL5295.C3	Synaptotagmin-like protein 5	45.449	21	0.06 ± 0.05	**0.44**	0.01
CL10768.C1	Choline transporter-like protein 2, partial	78.431	9	0.07 ± 0.01	7.38	3.24
Unigene434	Myosin-Id	104.954	6	0.05 ± 0.01	3.51	13.52
Protein degradation						
CL1290.C1	N-acetylated-alpha-linked acidic dipeptidase-like protein[c]	17.286	16	0.36 ± 0.18	4853.39	47.40
CL4546.C1	Cysteine proteases inhibitor	21.102	10	0.30 ± 0.30	138.64	338.14
CL14.C1	Leukocyte elastase inhibitor-like	40.886	12	0.19 ± 0.10	3.61	75.71
CL1656.C1	Amiloride-sensitive amine oxidase [copper-containing] [c]	84.207	47	0.17 ± 0.06	131.34	1.73
Unigene35321	Transmembrane protease, serine 2	25.376	20	0.10 ± 0.01	**115.88**	0.04
CL1819.C1	Transmembrane protease, serine 9	29.842	7	0.13 ± 0.06	**182.26**	0.37
CL7499.C1	alpha-1-inhibitor 3-like	71.366	9	0.08 ± 0.01	5.00	9.94
Energy metabolism						
Unigene6216	ADP-ribosylation factor 2-like	21.048	15	0.40 ± 0.19	30.90	31.66
CL9967.C1	15-hydroxyprostaglandin dehydrogenase [NAD+]	25.919	32	0.24 ± 0.10	52.09	0.84
CL10187.C1	4-hydroxyphenylpyruvate dioxygenase	43.541	17	0.18 ± 0.06	2.96	6.68
CL6535.C1	aldehyde dehydrogenase family 3 member B1-like isoform X1	54.884	24	0.14 ± 0.00	0.99	0.87
Transcription and translation						
CL2610.C3	apoptosis-inducing factor 3-like protein	58.429	108	0.43 ± 0.10	19.15	0.51
Unigene28387	ribosomal protein S14	16.193	37	0.11 ± 0.05	117.25	141.78
CL7340.C2	elongation factor 1 alpha	50.033	55	0.09 ± 0.02	1263.38	1555.68
Unigene18923	40S ribosomal protein S6	14.733	12	0.13 ± 0.08	1.28	0.89
Others						
Unigene34633	BRAFLDRAFT_214672	15.939	38	1.55 ± 0.68	2351.89	0.63
CL2247.C1	Novel protein	29.305	130	1.20 ± 0.79	**184.50**	0.11
Unigene40414	LOTGIDRAFT_234762[c]	24.551	33	0.73 ± 0.38	**1172.28**	0.12
Unigene28826	Lustrin A, partial	16.64	34	0.46 ± 0.12	593.08	64.75
Unigene34623	Novel protein[c]	34.624	34	0.35 ± 0.12	**1537.44**	0.00
Unigene6438	BRAFLDRAFT_113974	53.817	49	0.25 ± 0.05	138.29	26.28

Table 1 (*continued*)

Name	Annotation	MW (KDa)	No. of peptides	Average %[a]	Pm_TPM[b]	
					AG	OT
CL677.C1	Tetraspanin, partial	25.157	12	0.10 ± 0.01	13.34	4.67
Unigene7330	CGI_10006272	47.604	4	0.09 ± 0.03	15.80	20.58
CL8811.C1	LOC101850990	122.888	27	0.02 ± 0.01	8.46	9.56

[a] Normalized emPAI values. Data are expressed as mean ± standard deviation based on two or three biological replicates.
[b] Transcriptomic expression levels of genes in albumen gland (AG) and other tissues (OT). Data are expressed in transcripts per million (TPM). Bold TPM values: Genes only expressed in AG (TPM < 0.5 in OT).
[c] Indicates that the protein has a predicted signal peptide.

subunit compositions [12]. Therefore, more work is needed to better understand the relationship between the differences in subunit abundance of this perivitellin fraction within and between the two species and the impact this may have in embryonic development.

Apart from the pigmented perivitellins that are responsible for the conspicuous reddish coloration, antidigestive and antinutritive defenses against predators [28], immune responsive proteins are a potentially important functional group in *P. maculata* eggs that protect the embryos from pathogens. Among them, the iron storage protein ferritin, the ninth most abundant PVF protein (~1.38%) in *P. maculata*, has been shown to actively participate in mollusk immunity [47,48]. It is a major exogenous yolk protein in the freshwater snails *Lymnaea stagnalis* and *Planorbarius corneus* [47]. Ferritin was overexpressed in *P. canaliculata* undergoing estivation, which might prevent iron binding to pathogens to reduce their growth [49]. Kunitz-like protease inhibitor also plays an essential role in innate immunity by serving as an inhibitor of serine proteinase in the razor clam *Solen grandis* [50]. Although its protein abundance in our study was relatively low (~0.33%), the transcript expression level in albumen gland (TPM = 1938.09) was much higher than that in other tissues (TPM = 0.30), indicating its high transcription rate in the AG of *P. maculata*. In fact *P. maculata* eggs display a moderate serine protease inhibition activity [12]. Serpin 1 and serine protease inhibitor 1 may protect the egg masses against foreign serine protease released by bacteria and viruses. They may also function in melanization and immune response by activating phenoloxidase [51,52]. Several other immune-related proteins (e.g., C1q domain-containing protein, peptidoglycan recognition protein S1L and thioester-containing protein-C) are also present in the PVF of *P. maculata*, and these proteins have been detected in the PVF of *P. canaliculata* [10] and mucosal secretions of *Crassostrea virginica* [53]. The transcript abundances of C1q domain-containing protein and peptidoglycan recognition protein S1L in AG (TPM = 19.602.56 and 51.85, respectively) are much higher than those in OT (TPM = 0.06 and 0.00, respectively), indicating that these proteins may be specifically secreted for packaging into the egg.

A recent study on the newly laid eggs of *P. maculata* shows that the PVF contains 76.4% carbohydrates, 18.7% proteins, and 1.2% lipid in dry weight [12], which is in the same range observed for its sister species *P. canaliculata* [11]. These molecules serve as an energy and carbon sources during the embryonic development, which requires a wide variety of enzymes. Our study detected several carbohydrate, protein, and lipid metabolizing enzymes in *P. maculata* PVF, which are readily available even before the embryo develops and starts to secret the other enzymes required to fully metabolize the PVF. The PVF enzymes include 4-hydroxyphenylpyruvate dioxygenase (HPPD), serine and cysteine protease inhibitors and 15-hydroxyprostaglandin dehydrogenase (NAD+). HPPD, a Fe(II)-dependent, non-heme oxygenase, functions in tyrosine catabolic pathway which is common to almost all aerobic life forms [54]. Cysteine protease inhibitor plays various roles in physiological and pathological processes such as inhibiting protein processing and degradation as well as MHC class II immune responses [55,56]. 15-hydroxyprostaglandin dehydrogenase (NAD+) has alcohol dehydrogenase activity and is involved in the metabolism of lipid signaling molecules. For example, some glycolysis-related enzymes (e.g. dihydrolipoamide *S*-acetyltransferase, pyruvate dehydrogenase β, enolase, phosphoglycerate kinase 1 and glyceraldehyde 3-phosphate dehydrogenase) have been reported to play a critical role in the embryonic development of the snails *P. canaliculata* and *Lymnaea stagnalis* [57,58].

4.2. Evolutionary characterization of PVF proteins

A self-BLAST of the *P. maculata* PVF protein coding genes with an *E*-value threshold of $1 \times e^{-5}$ revealed 29 pairs of homologs, indicating these proteins have undergone several duplication even`ts during their evolutionary history. However, similar to the PVF genes of *P. canaliculata* [10], the sequence similarities between the PmPVF paralogs

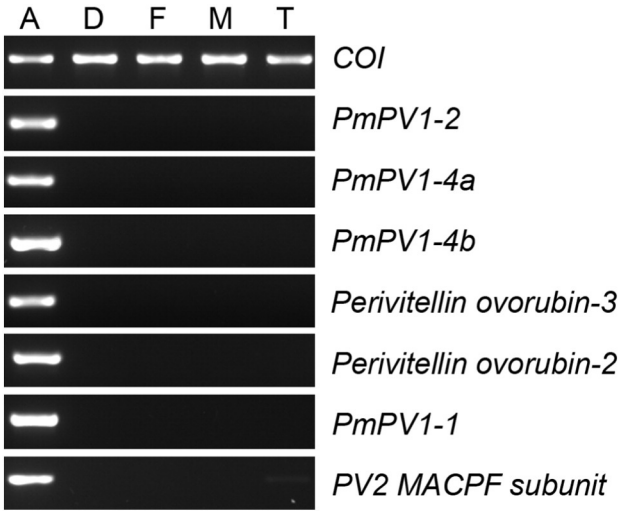

Fig. 2. Gene expression analysis of seven highly-abundant PVF protein-coding genes and a cytochrome *c* oxidase I (COI) gene. The capital letters on top of each lane represent different organs. A: albumen gland; D: digestive gland; F: foot; M: mantle; T: testis.

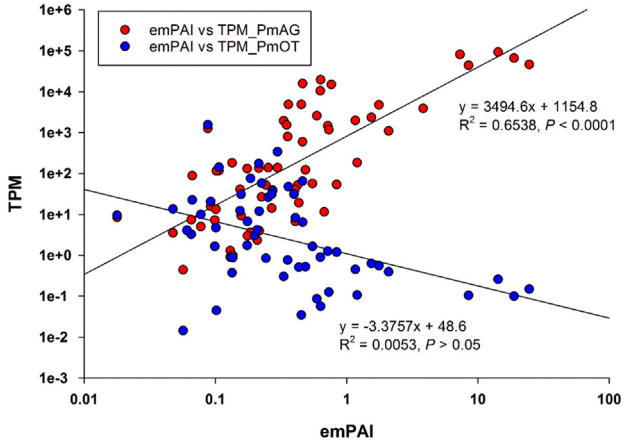

Fig. 3. Correlation between the abundance (emPAI) of 64 PVF proteins and their corresponding gene expression levels (TPM) in albumen gland (AG) and other tissues (OT). The data were log-transformed.

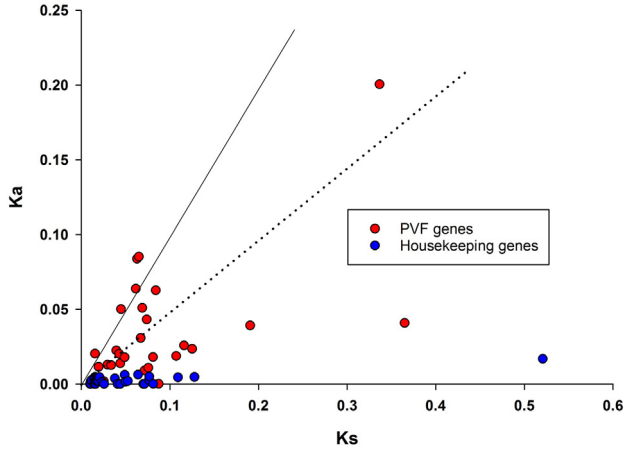

Fig. 4. Distribution of Ka, Ks and Ka/Ks values in 30 PVF protein coding genes and 30 housekeeping genes. The solid line and dotted line represents Ka/Ks = 1 and Ka/Ks = 0.5, respectively.

vary greatly from 27.1% to 99.7%, indicating some of these gene duplications must have occurred early in the diversification of *Pomacea*. A phylogenetic analysis of PmPV1 paralogs and sequence alignment between the four PmPV1 paralog subunits in *P. maculata* and their corresponding orthologs in *P. canaliculata* revealed several fully conserved amino acid sites and motifs such as the IXGGP motif.

Since PVF proteins in ampullariids play vital roles and their differential composition is expected to be critical for the acquisition of various oviposition strategies [6,8], we hypothesize that at least some of these proteins must be under high pressure of adaptive selection. Base substitution analysis of PVF protein orthologs between the sister species *P. maculata* and *P. canaliculata* indeed showed that their sequence divergence varied substantially. Among them, five PVF genes had a Ka/Ks value >1.0, a clear indication that they have undergone positive selection. In addition, six genes had a Ka/Ks value between 0.5 and 1, which are candidate genes for further examination to evaluate whether certain sequence regions or amino acid sites have undergone positive selection [19]. These 11 genes with relatively high Ka/Ks values encode not only protein subunits of the major multifunctional perivitellins, but also proteins involved in signaling and binding, immunity, protein degradation, and energy metabolism functions, indicating widespread positive selection in PVF protein coding genes in these freshwater snails adopting aerial oviposition. Comparing the Ka, Ks, and Ka/Ks values between PVF protein-coding genes with a set of housekeeping genes shows that the PVF genes have significantly higher mean Ka, Ks and Ka/Ks values than the housekeeping genes. This result indicates that the PVF genes are under weaker selective constraints than the housekeeping genes, further supporting previous studies showing that reproductive proteins evolve faster than housekeeping genes in various animals [19,59]. Since the Ks values are quite similar between PVF and housekeeping genes, the higher evolutionary rate of PVF protein-coding genes is likely due to the higher rates of nonsynonymous substitutions.

4.3. Possible functional consequences of base substitution in some PVF proteins

Homology modeling shows that, in some of PVF proteins, the nonsynonymous substitutions are located at or near the potential binding regions, suggesting the mutation may modify protein conformation and thus protein function. For example, ferritin, a protein that maintains iron homeostasis and provides antioxidant and antimicrobial protection have several nonsynonymous substitution sites in this apple snail. Of interest are the 8E located near the ferrihydrite nucleation center (6D and 7E), and 65K and 66I located near an iron ion channel (64H). Mutations at these sites could change the conformation of the 3D structure, and thus weaken or strengthen the iron transport and storage functions [60,61].

5. Conclusions

The present study provides the first comprehensive proteomic analysis of *P. maculata* egg perivitelline fluid and the development of a valuable transcriptomic resource for a non-model organism which is an invasive species and serious crop pest. Albumen gland exclusively expresses many of the PVF proteins, highlighting the critical role of this accessory gland in apple snail reproduction. The maternal investment is not limited to nutrition and protection as previously reported for the major perivitellins. Evolutionary analysis between *P. maculata* and the sister species *P. canaliculata* proteins shows that around 8% of the PVF protein-coding genes have undergone positive selection. Some of these proteins have experienced nonsynonymous amino acid substitutions at or near the ligand binding sites, potentially affecting protein conformation and function. Overall, our study provides the proteomic and transcriptomic foundation for further functional and evolutionary analysis of reproductive proteins in a group of snails that have shifted from aquatic to aerial oviposition, a strategy that has seldom occurred in the animal kingdom.

Fig. 5. Predicted 3D structure (A) and amino acid sequence (B) of *P. maculata* ferritin (partial). The locations of potential ligand binding sites (yellow), nonsynonymous substitution sites (red) and overlapped ligand binding and nonsynonymous substitution sites (green) are highlighted.

Abbreviations

AG	albumen gland
CID	collision-induced dissociation
COI	cytochrome *c* oxidase subunit I
emPAI	Exponentially Modified Protein Abundance Index
HCD	high-energy collision-induced dissociation
HPPD	4-hydroxyphenylpyruvate dioxygenase
Ka	nonsynonymous substitution rate
Ks	synonymous substitutions per synonymous site
LC-MS/MS	Liquid chromatography-mass spectrometry/mass spectrometry
OT	other tissues
PVF	perivitelline fluid
RPM	reads per kilobase
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TPM	transcripts per million

Data accessibility

Protein sequences translated from the combined transcriptomes of albumen gland and other tissues can be found in a Data in Brief paper (Mu et al., submitted) [62].

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

This project was supported by the Research Grants Council (grant no. HKBU 12301415) and Hong Kong Baptist University (grant no. FRG1/14-15/026), China and Agencia Nacional de Promoción Científica y Tecnológica, Argentina (grant No. 0850). We thank Haiwei Luo from The Chinese University of Hong Kong for helpful discussion on evolutionary analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2017.01.006.

References

[1] W.J. Swanson, V.D. Vacquier, The rapid evolution of reproductive proteins, Nat. Rev. Genet. 3 (2002) 137–144.
[2] L. Shu, M.J.F. Suter, K. Räsänen, Evolution of egg coats: linking molecular biology and ecology, Mol. Ecol. 24 (2015) 4052–4073.
[3] D.D. Thomas, C.A. Donnelly, R.J. Wood, L.S. Alphey, Insect population control using a dominant, repressible, lethal genetic system, Science 287 (2000) 2474–2476.
[4] N.L. Clark, J.E. Aagaard, W.J. Swanson, Evolution of reproductive proteins from animals and plants, Reproduction 131 (2006) 11–22.
[5] K.A. Hayes, R.L. Burks, A. Castro-Vazquez, P.C. Darby, H. Heras, P.R. Martín, J.-W. Qiu, S.C. Thiengo, I.A. Vega, T. Wada, Y. Yusa, S. Burela, P. Cadierno, J.A. Cueto, F.A. Dellagnola, M.S. Dreon, M.V. Frassa, M. Giraud-Billoud, M.S. Godoy, S. Ituarte, E. Koch, K. Matsukura, M.Y. Pasquevich, C. Rodriguez, L. Saveanu, M.E. Seuffert, E.E. Strong, J. Sun, N.E. Tamburi, M.J. Tiecher, R.L. Turner, P.L. Valentine-Darby, R.H. Cowie, Insights from an integrated view of the biology of apple snails (Caenogastropoda: Ampullariidae), Malacologia 58 (2015) 245–302.
[6] K.A. Hayes, R.H. Cowie, A. Jørgensen, R. Schultheiß, C. Albrecht, S.C. Thiengo, Molluscan models in evolutionary biology: apple snails (Gastropoda: Ampullariidae) as a system for addressing fundamental questions, Am. Malacol. Bull. 27 (2009) 47–58.
[7] Y. Yusa, Predation on eggs of the apple snail *Pomacea canaliculata* (Gastropoda: Ampullariidae) by the fire ant *Solenopsis geminata*, J. Molluscan Stud. 67 (2001) 275–279.
[8] H. Heras, M.S. Dreon, S. Ituarte, R.J. Pollero, Egg carotenoproteins in neotropical Ampullariidae (Gastropoda: Arquitaenioglossa), Comp. Biochem. Physiol. C 146 (2007) 158–167.
[9] M.S. Dreon, M.V. Frassa, M. Ceolín, S. Ituarte, J.-W. Qiu, J. Sun, P.E. Fernandez, H. Heras, Novel animal defenses against predation: a snail egg neurotoxin combining lectin and pore-forming chains that resembles plant defense and bacteria attack toxins, PLoS One 8 (2013), e63782. .
[10] J. Sun, H. Zhang, H. Wang, H. Heras, M.S. Dreon, S. Ituarte, T. Ravasi, P.-Y. Qian, J.-W. Qiu, First proteome of the egg perivitelline fluid of a freshwater gastropod with aerial oviposition, J. Proteome Res. 11 (2012) 4240–4248.
[11] H. Heras, C.F. Garin, R.J. Pollero, Biochemical composition and energy sources during embryo development and in early juveniles of the snail *Pomacea canaliculata* (Mollusca: Gastropoda), J. Exp. Zool. 280 (1998) 375–383.
[12] M.L. Giglio, S. Ituarte, M.Y. Pasquevich, H. Heras, The eggs of the apple snail *Pomacea maculata* are defended by indigestible polysaccharides and toxic proteins, Can. J. Zool. (2017)http://dx.doi.org/10.1139/cjz-2016-0049 (in press).
[13] M.S. Dreon, H. Heras, R.J. Pollero, Characterization of the major egg glycolipoproteins from the perivitellin fluid of the apple snail *Pomacea canaliculata*, Mol. Reprod. Dev. 68 (2004) 359–364.
[14] H. Heras, M.V. Frassa, P.E. Fernandez, C.M. Galosi, E.J. Gimeno, M.S. Dreon, First egg protein with a neurotoxic effect on mice, Toxicon 52 (2008) 481–488.
[15] M.S. Dreon, S. Ituarte, H. Heras, The role of the proteinase inhibitor ovorubin in apple snail eggs resembles plant embryo defense against predation, PLoS One 5 (2010), e15059.
[16] M.V. Frassa, M. Ceolín, M.S. Dreon, H. Heras, Structure and stability of the neurotoxin PV2 from the eggs of the apple snail *Pomacea canaliculata*. Biochimica et Biophysica Acta (BBA)-Proteins and, Proteomics 1804 (2010) 1492–1499.
[17] S. Ituarte, M.S. Dreon, M.Y. Pasquevich, P.E. Fernández, H. Heras, Carbohydrates and glycoforms of the major egg perivitellins from *Pomacea* apple snails (Architaenioglossa: Ampullariidae), Comp. Biochem. Physiol. B 157 (2010) 66–72.
[18] M.S. Dreon, P.E. Fernández, E.J. Gimeno, H. Heras, Insights into embryo defenses of the invasive apple snail *Pomacea canaliculata*: egg mass ingestion affects rat intestine morphology and growth, PLoS Negl. Trop. Dis. 8 (2014), e2961.
[19] W.J. Swanson, A.G. Clark, H.M. Waldrip-Dail, M.F. Wolfner, C.F. Aquadro, Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 7375–7379.
[20] M. Schein, Z. Yang, T. Mitchell-Olds, K.J. Schmid, Rapid evolution of a pollen-specific oleosin-like gene family from *Arabidopsis thaliana* and closely related species, Mol. Biol. Evol. 21 (2004) 659–669.
[21] G.D. Findlay, W.J. Swanson, Proteomics enhances evolutionary and functional analysis of reproductive proteins, BioEssays 32 (2010) 26–36.
[22] K.A. Hayes, R.H. Cowie, S.C. Thiengo, E.E. Strong, Comparing apples with apples: clarifying the identities of two highly invasive Neotropical Ampullariidae (Caenogastropoda), Zool. J. Linnean Soc. 166 (2012) 723–753.
[23] T.A. Rawlings, K.A. Hayes, R.H. Cowie, T.M. Collins, The identity, distribution, and impacts of non-native apple snails in the continental United States, BMC Evol. Biol. 7 (2007) 1.
[24] K.A. Hayes, R.C. Joshi, S.C. Thiengo, R.H. Cowie, Out of South America: multiple origins of non-native apple snails in Asia, Divers. Distrib. 14 (2008) 701–712.
[25] K.L. Kwong, P.K. Wong, S.S.S. Lau, J.-W. Qiu, Determinants of the distribution of apple snails in Hong Kong two decades after their initial invasion, Malacologia 50 (2008) 293–302.
[26] S. Lv, Y. Zhang, H.X. Liu, L. Hu, Q. Liu, F.R. Wei, Y.-H. Guo, W. Hu, X.-N. Zhou, J. Utzinger, Phylogenetic evidence for multiple and secondary introductions of invasive snails: *Pomacea* species in the People's Republic of China, Divers. Distrib. 19 (2013) 147–156.
[27] K. Matsukura, M. Okuda, N.J. Cazzaniga, T. Wada, Genetic exchange between two freshwater apple snails, *Pomacea canaliculata* and *Pomacea maculata* invading East and Southeast Asia, Biol. Invasions 15 (2013) 2039–2048.
[28] M.Y. Pasquevich, M.S. Dreon, H. Heras, The major egg reserve protein from the invasive apple snail *Pomacea maculata* is a complex carotenoprotein related to those of *Pomacea canaliculata* and *Pomacea scalaris*, Comp. Biochem. Physiol. B 169 (2014) 63–71.
[29] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome, Nat. Biotechnol. 29 (2011) 644–652.
[30] W.K.F. Tse, J. Sun, H. Zhang, A.Y.S. Law, B.H.Y. Yeung, S.C. Chou, J.-W. Qiu, C.K.C. Wong, Transcriptomic and iTRAQ proteomic approaches reveal novel short-term hyperosmotic stress responsive proteins in the gill of the Japanese eel (*Anguilla japonica*), J. Proteome 89 (2013) 81–94.
[31] D.B. Friedman, Quantitative proteomics for two-dimensional gels using difference gel electrophoresis (DIGE) technology, in: R. Matthiesen (Ed.), Mass Spectrometry Data Analysis in Proteomics, Humana Press, New Jersey 2007, pp. 219–239.
[32] H. Mu, J. Sun, L. Fang, T. Luan, G.A. Williams, S.G. Cheung, C.K.C. Wong, J.-W. Qiu, Genetic basis of differential heat resistance between two species of congeneric freshwater snails: insights from quantitative proteomics and base substitution rate analysis, J. Proteome Res. 14 (2015) 4296–4308.
[33] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinf. 12 (2011) 1.
[34] N.N. Rosic, M. Pernice, M. Rodriguez-Lanetty, O. Hoegh-Guldberg, Validation of housekeeping genes for gene expression studies in *Symbiodinium* exposed to thermal and light stress, Mar. Biotechnol. 13 (2011) 355–365.
[35] J. Sun, M. Wang, H. Wang, H. Zhang, X. Zhang, V. Thiyagarajan, P.-Y. Qian, J.-W. Qiu, *De novo* assembly of the transcriptome of an invasive snail and its multiple ecological applications, Mol. Ecol. Resour. 12 (2012) 1133–1144.
[36] Z. Zhang, J. Xiao, J. Wu, H. Zhang, G. Liu, X. Wang, L. Dai, ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments, Biochem. Biophys. Res. Commun. 419 (2012) 779–781.
[37] Z. Zhang, J. Li, X.Q. Zhao, J. Wang, G.K.-S. Wong, J. Yu, KaKs_Calculator: calculating Ka and Ks through model selection and model averaging, Genomics Proteomics Bioinformatics 4 (2006) 259–263.

[38] L.Y. Chen, S.Y. Zhao, Q.F. Wang, M.L. Moody, Transcriptome sequencing of three *Ranunculus* species (Ranunculaceae) reveals candidate genes in adaptation from terrestrial to aquatic habitats, Sci. Rep. 5 (2014) 10098.
[39] J.L. Villanueva-Cañas, S. Laurie, M.M. Albà, Improving genome-wide scans of positive selection by using protein isoforms of similar length, Genome Biol. Evol. 5 (2013) 457–467.
[40] L. Zhang, W.H. Li, Mammalian housekeeping genes evolve more slowly than tissue-specific genes, Mol. Biol. Evol. 21 (2004) 236–239.
[41] W. Lv, J. Zheng, M. Luan, M. Shi, H. Zhu, M. Zhang, H. Lv, Z. Shang, L. Duan, R. Zhang, Y. Jiang, Comparing the evolutionary conservation between human essential genes, human orthologs of mouse essential genes and human housekeeping genes, Brief. Bioinform. 16 (2015) 922–931.
[42] Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, M. Mann, Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein, Mol. Cell. Proteomics 4 (2005) 1265–1272.
[43] T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, Nat. Methods 8 (2011) 785–786.
[44] L.A. Kelley, S. Mezulis, C.M. Yates, M.N. Wass, M.J.E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis, Nat. Protoc. 10 (2015) 845–858.
[45] M.N. Wass, L.A. Kelley, M.J.E. Sternberg, 3DLigandSite: predicting ligand-binding sites using similar structures, Nucleic Acids Res. (2010) (gkq406).
[46] A. Marchler-Bauer, M.K. Derbyshire, N.R. Gonzales, S. Lu, F. Chitsaz, L.Y. Geer, R.C. Geer, J. He, M. Gwadz, D.I. Hurwitz, C.J. Lanczycki, F. Lu, G.H. Marchler, J.S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, S.H. Bryant, CDD: NCBI's conserved domain database, Nucleic Acids Res. (2014) (gku1221).
[47] W. Bottke, M. Burschyk, J. Volmer, On the origin of the yolk protein ferritin in snails, Roux's Arch. Dev. Biol. 197 (1988) 377–382.
[48] A.E. Lockyer, A.M. Emery, R.A. Kane, A.J. Walker, C.D. Mayer, G. Mitta, C. Coustau, C.M. Adema, B. Hanelt, D. Rollinson, L.R. Noble, C.R. Jones, Early differential gene expression in haemocytes from resistant and susceptible *Biomphalaria glabrata* strains in response to *Schistosoma mansoni*, PLoS One 7 (2012), e51102. .
[49] J. Sun, H. Mu, H. Zhang, K.H. Chandramouli, P.-Y. Qian, C.K.C. Wong, J.-W. Qiu, Understanding the regulation of estivation in a freshwater snail through iTRAQ-based comparative proteomics, J. Proteome Res. 12 (2013) 5271–5280.
[50] X. Wei, J. Yang, J. Yang, X. Liu, M. Liu, D. Yang, J. Xu, X. Hu, A four-domain Kunitz-type proteinase inhibitor from *Solen grandis* is implicated in immune response, Fish Shellfish Immunol. 33 (2012) 1276–1284.
[51] J.J.M. Hathaway, C.M. Adema, B.A. Stout, C.D. Mobarak, E.S. Loker, Identification of protein components of egg masses indicates parental investment in immunoprotection of offspring by *Biomphalaria glabrata* (Gastropoda, Mollusca), Dev. Comp. Immunol. 34 (2010) 425–435.
[52] I. González-Santoyo, A. Córdoba-Aguilar, Phenoloxidase: a key component of the insect immune system, Entomol. Exp. Appl. 142 (2012) 1–16.
[53] E.P. Espinosa, A. Koller, B. Allam, Proteomic characterization of mucosal secretions in the eastern oyster, *Crassostrea virginica*, J. Proteome 132 (2016) 63–76.
[54] G.R. Moran, 4-Hydroxyphenylpyruvate dioxygenase, Arch. Biochem. Biophys. 433 (2005) 117–128.
[55] P.B. Armstrong, Proteases and protease inhibitors: a balance of activities in host–pathogen interaction, Immunobiology 211 (2006) 263–281.
[56] R. Vicik, M. Busemann, K. Baumann, T. Schirmeister, Inhibitors of cysteine proteases, Curr. Top. Med. Chem. 6 (2006) 331–353.
[57] E.M. Goudsmit, Carbohydrates and carbohydrate metabolism in mollusca, in: M. Florkin, B.T. Scheer (Eds.), Chemical Zoology, Mollusca, vol. 7, Academic Press, New York 1972, pp. 219–244.
[58] J. Sun, Y. Zhang, V. Thiyagarajan, P.Y. Qian, J.-W. Qiu, Protein expression during the embryonic development of a gastropod, Proteomics 10 (2010) 2701–2711.
[59] J.A. Andrés, L.S. Maroja, S.M. Bogdanowicz, W.J. Swanson, R.G. Harrison, Molecular evolution of seminal proteins in field crickets, Mol. Biol. Evol. 23 (2001) 1574–1584.
[60] N.D. Chasteen, P.M. Harrison, Mineralization in ferritin: an efficient means of iron storage, J. Struct. Biol. 126 (1999) 182–194.
[61] T. Tosha, R.K. Behera, E.C. Theil, Ferritin ion channel disorder inhibits Fe (II)/O2 reactivity at distant sites, Inorg. Chem. 51 (2012) 11406–11411.
[62] H. Mu, J. Sun, H. Heras, K.H. Chu, J.-W. Qiu, Dataset for the proteomic and transcriptomic analyses of perivitelline fluid proteins in *Pomacea* snail eggs, Data in Brief (2017) (submitted for publication).